

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS FÍSICAS



TESIS DOCTORAL

**Un formalismo para representación y utilización de
información bilingüe en sistemas de traducción automática**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR
PRESENTADA POR

Pilar Rodríguez Marin

DIRECTOR:

Luis de Sopena Pastor

Madrid, 2015

UNIVERSIDAD COMPLUTENSE DE MADRID

Facultad de Ciencias Físicas

Departamento de Informática y Automática



UNIVERSIDAD COMPLUTENSE



5324959519

IT
UCM
1991

**UN FORMALISMO PARA LA
REPRESENTACION Y UTILIZACION DE
INFORMACION BILINGÜE
EN SISTEMAS DE TRADUCCION
AUTOMATICA**



b 1494523x
i 39767243

Pilar Rodríguez Marín
Madrid, 1991

Colección Tesis Doctorales. N.º 301/91

© Pilar Rodríguez Marín

**Edita e imprime la Editorial de la Universidad
Complutense de Madrid. Servicio de Reprografía.
Escuela de Estomatología. Ciudad Universitaria.
Madrid, 1991.
Ricoh 3700
Depósito Legal: M-42058 -1991**

**Un Formalismo para la
Representación y Utilización de Información Bilingüe
en Sistemas de Traducción Automática**

Junio 1990

Autor: Pilar Rodríguez Marín

Director: Luis de Sopeña Pastor

**Departamento de Informática y Automática
Facultad de Ciencias Físicas
Universidad Complutense de Madrid**

Memoria presentada para optar al grado de Doctor en Ciencias Físicas

Agradecimientos

Deseo expresar mi sincero agradecimiento en primer lugar a Luis de Sopeña, por haber aceptado ser director de esta tesis, así como por sus valiosos consejos y aliento constante. De igual modo quiero agradecer a José Jaime Ruz haber aceptado ser mi ponente, sin su apoyo y su ayuda esta tesis difícilmente hubiese llegado a buen término.

También quiero agradecer su gran ayuda a los miembros del grupo de traducción automática del Centro de Investigación UAM-IBM: Isabel Zapata y Teófilo Redondo, inestimable compañero de despacho. Sin su labor técnica y apoyo no hubiera sido posible realizar este trabajo. Gracias también a Celia Villar, Chelo Rodríguez y Carmen Valladares, cuya experiencia en muchos temas ha sido fundamental.

Y, muy especialmente, gracias a Ramón Casajuana, que ha soportado con infinita paciencia mis problemas con la estética de esta memoria.

Mis agradecimientos finales a todos mis compañeros del Centro de Investigación y del Instituto de Ingeniería del Conocimiento por el tiempo y los medios que me han permitido dedicar a la elaboración de este trabajo.

Contenido

Presentación	xv
---------------------------	-----------

Notas	xix
--------------------	------------

Capítulo 1: La Información Léxica en los Sistemas de Traducción Automática . . .	1
---	----------

1.1 Diccionarios Electrónicos	3
--	----------

1.1.1 Características	3
-----------------------------	---

1.1.2 Codificación Automática	8
-------------------------------------	---

1.2 Traducción Automática y Diccionarios	13
---	-----------

1.2.1 Modelo de Transferencia	18
-------------------------------------	----

1.2.2 Transferencia Léxica: Algunos Sistemas	21
--	----

1.2.2.1 ALPS TSS Transactive	22
------------------------------------	----

1.2.2.2 SYSTRAN	23
-----------------------	----

1.2.2.3 LOGOS	23
---------------------	----

1.2.2.4 METAL	24
---------------------	----

1.2.2.5 DLT	25
-------------------	----

1.2.2.6 EUROTRA	26
-----------------------	----

1.2.3 Recapitulación	27
----------------------------	----

Capítulo 2: Marco de la Investigación: el proyecto MENTOR	31
2.1 Proyecto MENTOR	33
2.2 Estructura del Sistema	35
2.2.1 El Analizador	39
2.2.2 La Transferencia	44
2.2.3 La Generación	50
2.2.3.1 Orden de los Complementos y Generación de Clíticos	51
2.2.3.2 Omisión de pronombres	52
2.2.3.3 Tratamiento de pasivas	52
2.2.3.4 Presentación final de la oración	53
2.2.4 Los Diccionarios	53
2.2.4.1 Diccionario de análisis	54
2.2.4.2 Diccionario de generación	55
2.2.4.3 Diccionario de transferencia	55
2.3 Transferencia Léxica para una palabra: el verbo MAKE	59
2.3.1 Información disponible	60
2.3.2 Extracción de la Información	61
2.3.2.1 Acepciones Transitivas	64
2.3.2.2 Acepciones Intransitivas	70
2.3.3 Conclusiones sobre el ejemplo	74

Capítulo 3: Un Formalismo de Diccionario Bilingüe	77
3.1 Representación de la Información	79
3.1.1 Necesidades Lingüísticas	79
3.1.2 Solución computacional propuesta	81
3.2 Símbolos de Control	85
3.3 Primitivas	93
3.3.1 Decisiones	94
3.3.1.1 Decisiones de tipo C	97
C-ASPECT	97
C-TYPE	99
3.3.1.2 Decisiones de tipo F	100
F-CC	100
F-INDOBJ	101
F-IS	101
F-MODHEAD	103
F-NOFADJ	103
F-NOFPP	104
F-OBJ	105
F-SUBJ	107
3.3.1.3 Decisiones de tipo P	108
P-AJP	108
P-NP	109
P-PRMODS	109
P-PRPTCL	110
P-PSMODS	111

P-PTC	113
P-TYPE	114
3.3.1.4 Decisiones de tipo M	115
M-FEAT	115
M-IS	116
M-MORPH	116
M-SEG	117
3.3.2 Acciones	118
3.3.2.1 Acciones de tipo G	119
G-CC	119
G-FEAT	120
G-FLEX	122
G-SETPREP	123
G-STATE	124
3.3.2.2 Acciones de tipo S	124
S-COMPCL	124
S-FCHANGE	125
S-INFNP	126
S-RELCL	127
S-TBASES	128
S-XCHANGE	128
3.4 Almacenamiento y Acceso	131
3.4.1 Almacenamiento del Diccionario	131
3.4.2 Acceso	132
3.4.2.1 Actualización y Consulta	132
3.4.2.2 Acceso por programa	134

Capítulo 4: Aplicación del formalismo en MENTOR	137
4.1 Ejemplos de representación	139
4.1.1 RETURN: nombre, adjetivo y verbo	139
4.1.2 MAKE: verbo	141
4.2 Desarrollo de la traducción de una oración	143
4.3 Análisis de algunos casos con el verbo MAKE	161
Ejemplo 3: "This table is made of wood"	163
Ejemplo 6: "You should make the previous function a new tool"	164
Ejemplos 7/8: "They will make her/him president"	166
Ejemplo 11: "The circumstances made him write a new book"	167
Ejemplo 12: "This collection of functions has been made into a new program"	168
Ejemplo 17: "The car is only made up of tubes and wheels"	169
Ejemplo 18: "You can't make away easily"	171
4.4 Entradas y Primitivas: Evolución	173
<hr/>	
Conclusiones	177
<hr/>	
Referencias	181

- x -

Apéndices	191
Apéndice A. Notación	195
Apéndice B. Marcas Semánticas	199
Apéndice C. Algunos datos de los programas	207
Apéndice D. Algunas entradas completas	209

Figuras

1. Clasificación de los Diccionarios Electrónicos	6
2. Codificación de la Información Léxica	8
3. Modelo de transferencia	19
4. MENTOR: análisis	37
5. MENTOR: preprocesamiento	37
6. MENTOR: transferencia	37
7. MENTOR: generación	38
8. Árbol de análisis proporcionado por PEG	40
9. Ejemplo de un registro de la estructura: el verbo	40
10. Ejemplo de un registro de la estructura: la oración	42
11. Representación al comienzo de la fase de transferencia	43
12. Representación al comienzo de la fase de transferencia (cont.)	44
13. Ejemplo de detección de funciones profundas en la oración	46
14. Relación entre las transferencias léxica y estructural	48
15. Ejemplo de análisis ambiguos	49
16. Duplicación y reordenación de pronombres	52
17. Algunas entradas en el diccionario de análisis OD	54
18. Definiciones en distintos diccionarios	61
19. Formalización de acepciones (1)	65
20. Formalización de acepciones (2)	67
21. Formalización de acepciones (3)	67
22. Formalización de acepciones (4)	69

23. Formalización de acepciones (5)	70
24. Formalización de acepciones (6)	71
25. Formalización de acepciones (7)	72
26. Formalización de acepciones (8)	72
27. Formalización de acepciones (9)	73
28. Formalización de acepciones (10)	74
29. Formalización de acepciones: resultado final	75
30. Condiciones sobre la lengua origen	81
31. Símbolos de control	85
32. Aspecto: formas verbales.	98
33. Acceso al diccionario bilingüe a través de la interfaz	132
34. Entrada para el verbo MAKE: acepciones	141
35. Árbol de análisis proporcionado por PEG	143
36. Registro DECL1	144
37. Registro VP1	145
38. Registro VERB1	146
39. Registro VERB2	146
40. Representación al comienzo de la fase de transferencia	148
41. Representación al comienzo de la fase de transferencia (cont.)	149
42. Asignación de la traducción de "allow"	150
43. Transferencia Léxica (1)	151
44. Transferencia Léxica: resultado final (2)	158
45. Regencia en la oración	157
46. Tiempos de ejecución en ms: LISP interpretado	158
47. Tiempos de ejecución en ms: LISP Compilado	159

48.	Ejemplo 6: árboles de análisis proporcionados por PEG	165
49.	Ejemplos 7/8: árboles de análisis proporcionados por PEG	167
50.	MAKE: análisis ambiguo	168
51.	MAKE frasal: indicación en los registros	170
52.	MAKE frasal: aproximación	171
53.	Resultados sobre las entradas	174
54.	Primitivas Utilizadas	175
55.	Marcas semánticas para nombres	199
56.	Marcas semánticas para adjetivos	202
57.	Marcas semánticas de PEG para verbos	204
58.	Marcas semánticas para verbos	204

Presentación

La Traducción Automática (TA) constituye un área de investigación de gran interés [Sop88, Sop89]. Por una parte representa en sí una importante aplicación del Procesamiento del Lenguaje Natural (PLN), y por la otra, proporciona un marco excelente para la investigación en otros campos relacionados, ya que la implementación de un sistema de TA implica la reunión coherente de teorías sobre distintos temas: desde estudios gramaticales hasta estructuras de representación del conocimiento. Por lo tanto, puede servir para comprobar la validez práctica de estas mismas teorías, en el sentido de su aplicabilidad real.

Desde el punto de vista comercial y de desarrollo es también una aplicación importante por derecho propio: la traducción de documentos (legales, técnicos, etc.) es imprescindible para cualquier organización con intereses internacionales y, cada vez más, su mecanización llega a convertirse en una necesidad imperiosa.

Además, las coordenadas socio-económicas actuales evolucionan cada vez con más rapidez hacia situaciones plurilingües, creciendo la demanda de traducciones. Baste analizar el caso de una institución como la Comunidad Económica Europea (CEE), en la que existen nueve idiomas oficiales (alemán, danés, español, francés, griego, holandés, inglés, italiano y portugués), todos ellos con el mismo valor legal. Como consecuencia de esta situación, la CEE está gastando casi la mitad de su presupuesto de administración en traducción e interpretación [Bal88], y, aún así, muchos textos quedan sin traducir, por falta de tiempo o bien por falta de personal adecuado, ya que hay que tener en cuenta que en muchos casos se trata de textos técnicos que utilizan un lenguaje muy específico, por lo que su traducción requiere una formación altamente especializada y un esfuerzo considerable.

Una situación parecida se presenta en la ONU, donde el coste en estas materias se eleva a unos 180 millones de dólares al año, o en Canadá, con dos idiomas oficiales (inglés y francés) a los que han de traducirse todos los documentos. Similar es el problema en las comunidades autónomas bilingües españolas, aunque en ellas el volumen de material por traducir es menor.

A pesar de lo anterior, no son los organismos oficiales los mayores consumidores de traducciones, sino las empresas comerciales, que traducen más de las tres cuartas partes del total mundial [Van83]. La superación de las fronteras nacionales, en cuanto a intercambios comerciales se refiere, hace que, si de verdad se quiere competir en el mercado, sea necesario traducir la documentación que acompaña a cualquier producto. Pensemos, sin ir más lejos, en los prospectos de los medicamentos, las instrucciones de cualquier electrodoméstico, los manuales de productos informáticos o las especificaciones de una instalación.

Para hacernos una idea del volumen de trabajo de que estamos hablando, consideremos que el número medio de páginas que traduce un traductor profesional al día es de cinco, y que, por ejemplo, los manuales del sistema operativo UNIX de ATT tienen unas 2.400 páginas; esto significa

que una persona tardaría unos dos años en traducirlos, y, si se pretende disponer de la traducción en un periodo corto, digamos un mes, sería necesario el trabajo conjunto de 24 traductores.

Por otra parte, además del gasto real que supone la traducción, existen importantes consecuencias laterales, ya que se produce un retraso prácticamente inevitable en la distribución de los productos, con las pérdidas económicas que ello implica. Además, no se puede decir que la necesidad de realizar traducciones haya alcanzado una situación estable, ya que el crecimiento de la demanda estimado en el futuro, a nivel mundial, es del orden de un 15 a un 25% por año.

Esta situación justifica el renovado interés por la TA que se detecta a nivel mundial en centros de investigación y empresas, abordando el problema con un planteamiento mucho más realista que el que se tenía en sus comienzos, hace ya cuatro décadas. Para una exposición detallada de la historia, situación actual y perspectivas de la TA, nos remitimos al completísimo artículo de J. Slocum en *Computational Linguistics* [Slo85].

Aunque es cierto que hoy se sigue cuestionando la posibilidad de que una computadora lleve a cabo automáticamente una traducción (recordemos, por ejemplo, el artículo de M. Kay "*Machine Translation will not work*" [Kay86]) esta controversia posiblemente se reduce a ponerse de acuerdo sobre qué se entiende por TA. Evidentemente, conseguir traducir cualquier tipo de texto, de forma automática, y con la calidad necesaria para su publicación, es todavía impensable; pero, para que un sistema de TA sea útil no es necesario exigir tanto. Basta con que su uso aumente de forma significativa el rendimiento del traductor, la cantidad de material traducido en un tiempo dado. De hecho, muchos de los sistemas de TA hoy existentes necesitan de la intervención humana en alguno de los estadios del proceso: pre-editando los textos originales (para eliminar ambigüedades y construcciones problemáticas), interaccionado en mayor o menor medida con el traductor durante el proceso de traducción o post-editando los textos obtenidos (revisión del resultado por un traductor humano, fase de todos modos necesaria también tras un proceso manual de traducción).

De manera que el sentimiento generalizado es que la TA, entendida de forma realista, e incluyendo toda una serie de herramientas de ayuda a la traducción, es y será cada vez de más utilidad dentro de los ámbitos polilingües.

Son muchas, y de muy diversa índole, las técnicas lingüísticas e informáticas que hay que formalizar para poder afrontar el problema de la traducción automática: análisis morfológico, sintáctico y semántico, léxicos (monolingües, bilingües y multilingües), análisis contrastivo de pares de lenguas, generación de textos, etc. Cada uno de estas materias es de enorme interés, aunque no en todas se ha trabajado con la misma intensidad. Entre ellas se destaca, por sus características muy particulares, el estudio de los léxicos bilingües como estudio formal de la transferencia léxica entre los pares de lenguas: cómo expresar la relación entre los significantes y qué consecuencias tiene en la oración la elección de una traducción en concreto.

El tema de la transferencia léxica es enormemente complejo, y sólo puede abordarse a partir de una base lingüística muy sólida en la que se apoyen las formalizaciones propuestas [Melb88, Mel88b]. Incluye la determinación de las traducciones adecuadas para cada palabra en cualquiera de los posibles contextos en que pudiera aparecer, para lo cual es necesario determinar las condiciones que hacen distintas unas ocurrencias de otras. También es necesario indicar si se produce o no algún tipo de transformaciones en la oración por la elección de una traducción determinada. Y, si se quiere implementar un proceso de transferencia léxica automático, toda esta información ha de codificarse [Bog89], determinándose un mecanismo adecuado para que pueda ser interpretada por un programa.

En la presente memoria es éste el problema que se aborda, el de la *transferencia léxica* automática entre un par de lenguas: el inglés y el español, en el marco de un proyecto de investigación de traducción automática. Dicho proyecto se está llevando a cabo en el Centro de Investigación UAM-IBM, en colaboración con otros grupos y universidades.

El Capítulo 1, "*La Información Léxica en los Sistemas de Traducción Automática*", introduce el tema de la información léxica: sus características peculiares y cómo éste está presente en los sistemas actuales a través de los llamados *Diccionarios Electrónicos*; pasando después a relacionar este tipo de información con los sistemas de traducción automática, demostrando la importancia de la transferencia léxica en esta clase de programas, que pueden incluso estar basados principalmente en los diccionarios que utilizan. Se presentan también los tipos de diccionarios bilingües de que disponen algunos sistemas actuales, para terminar con una breve recapitulación sobre las características que éstos poseen.

En el Capítulo 2, "*Marco de la Investigación: el proyecto MENTOR*", se describe el prototipo de traducción automática que se ha implementado en el Centro de Investigación UAM-IBM: cómo ha sido diseñado y cuál es su estructura. Este conocimiento permite posteriormente hacer referencia a la relación entre los distintos módulos del sistema, en aquellos momentos en que ésta tiene interés. Por último, enlazando con el capítulo siguiente, se ha desarrollado un ejemplo de transferencia léxica para una palabra en concreto, a través del cual se ejemplifican los puntos críticos de este tipo de operaciones.

El Capítulo 3, "*Un formalismo de Diccionario Bilingüe*", expone ya el modelo de diccionario propuesto para la transferencia léxica; un modelo basado en la determinación de dos conjuntos de primitivas lingüísticas (en relación con las lenguas fuente y meta) y una gramática de definición, que incluye algunos símbolos de control. También describe los mecanismos de acceso y almacenamiento de la información léxica así definida.

En el Capítulo 4, "*Aplicación del formalismo en MENTOR*", se presenta el desarrollo completo de la traducción de una oración dentro del prototipo de traducción implementado, y que utiliza el formalismo de representación descrito anteriormente. Se exponen, asimismo, algunos casos particulares, que sirven para profundizar en el uso real y la potencia de formalismo presentado.

Se presentan finalmente las conclusiones a nuestro trabajo, y, tras las referencias, algunos apéndices que contienen información y datos de interés.

Notas

En esta memoria se ha utilizado la notación que habitualmente se emplea en la materia. La gran mayoría de los conceptos y palabras existen y se usan en español, pero hay otras que se ha preferido mantener en el original (inglés) por razones de claridad. Pedimos excusas por ello.

Para referirnos a los elementos gramaticales de la oración, hemos mantenido la notación que G. Heidorn y K. Jensen utilizan en la descripción de su gramática, PEG, ya que nos apoyamos en ésta [Jen86, Jen87]. Pensamos que cambiarla podría dificultar la exposición. Sin ser exhaustiva, la lista que presentamos a continuación contiene la mayoría de estos conceptos:

- NP (Noun Phrase): Sintagma Nominal
- VP (Verb Phrase): Sintagma Verbal
- PP (Prepositional Phrase): Sintagma Preposicional
- AVP (Adverb Phrase): Sintagma Adverbial
- AJP (Adjective Phrase): Sintagma Adjetivo
- INFCL (Infinitive Clause): Oración de Infinitivo
- THATCL (Completive or Relative Clause): Oración Completiva o Subordinada Relativa

**Capítulo 1: La Información Léxica en los Sistemas de
Traducción Automática**

1.1 Diccionarios Electrónicos

En este primer apartado se van a definir y caracterizar los diccionarios en cuanto éstos se constituyen en el principal instrumento de recopilación de la información léxica. A partir del establecimiento de los conceptos derivados de esta idea, se introduce posteriormente el de *Diccionario Electrónico*, haciendo hincapié en las similitudes y diferencias que presenta frente a los diccionarios que podríamos llamar clásicos, y en los principales problemas que surgen cuando el usuario final de estas herramientas es un programa y no un ser humano. Uno de los problemas más importantes es el de la necesidad de codificar la información léxica para que ésta se haga comprensible por el programa.

Para terminar, se considera la posibilidad de que la codificación de la información se realice de manera automática a partir de diccionarios expresados de la forma habitual.

1.1.1 Características

Bajo el término 'léxico' se incluyen varios conceptos diferentes, todos ellos íntimamente ligados a lo que es el *significado*. Así, se denomina *léxico* al conjunto de las palabras y expresiones utilizado habitualmente por una persona, o que se utiliza comúnmente en un cierto ámbito. También se denomina *léxico* al compendio de palabras, giros, etc., ordenados según un criterio determinado, generalmente alfabético. La información léxica, relativa a un conjunto de palabras determinado, se recoge en un *diccionario*, que Lázaro Carreter [Car84] describe como:

diccionario

Libro en que, por orden alfabético generalmente, se contienen y explican las significaciones de todas las palabras de una lengua, o se ponen en correspondencia con las equivalentes de otro u otros idiomas.

De acuerdo con la descripción anterior, que sólo sufre ligeras variaciones de unos autores a otros, la tarea de la *lexicografía*, como ciencia que cuyo objeto es el léxico, es describir todas o algunas de las palabras de uno o más lenguajes según unos criterios prefijados. Pero, a pesar de que el arte de hacer diccionarios es esencial y tan antiguo como las primeras comunicaciones escritas [AIK83], esto no significa que constituya un conocimiento completo y cerrado. De hecho, la creación de un nuevo diccionario obliga muchas veces a revisar y redefinir algunos conceptos fundamentales.

Aunque clásicamente se manejan y crean los diccionarios de la misma manera que los libros, utilizando el papel como soporte físico, la definición anterior casi inmediatamente sugiere la posibilidad de utilizar una computadora para realizar esa tarea: como elemento potente para procesar

información simbólica, almacenar gran cantidad de datos de manera óptima, recuperar estos datos de la forma requerida y codificar e interpretar la información dependiendo de ciertos criterios preestablecidos. Surge así lo que se ha dado en llamar Diccionario Electrónico (DE).

El término 'Diccionario Electrónico' (Electronic Dictionary), aunque quizá poco afortunado, se acuñó en respuesta a la necesidad de identificar de una forma específica a todo el conjunto de sistemas de información léxica que se implementan en una computadora; con almacenamiento magnético u óptico, sofisticados algoritmos de acceso y, principalmente, usos muy diversos. Es un término relativamente nuevo, ya que hasta hace muy poco tiempo este tipo de herramientas léxicas no ha alcanzado una entidad propia, considerándose como un componente más, en general secundario, de los sistemas de los que formaba parte. Sin embargo, gracias a las aplicaciones que se han desarrollado en los últimos tiempos, el componente léxico de los sistemas ha ido cobrando cada vez más importancia, pasando a ser básico en muchas y muy diversas aplicaciones, desde los programas de ayuda a la traducción en los sistemas de traducción asistida, hasta los sistemas de traducción completamente automática, pasando por las bases de datos terminológicas y de nomenclatura, los sistemas de análisis sintáctico y semántico, las bases de conocimiento, etc. cuyo soporte electrónico permite una mayor estandarización.

Este interés creciente por las aplicaciones léxicas en el ámbito informático se refleja en congresos y reuniones internacionales, p.e. el *International Symposium on Electronic Dictionaries*, y en el creciente número de proyectos de investigación y desarrollo que se están llevando a cabo actualmente en empresas y universidades.

Todas las aplicaciones informáticas relacionadas con el léxico se basan bien en la creación, bien en la utilización de diccionarios de uno u otro tipo, dependiendo de los objetivos. Es necesario, por lo tanto, analizar previamente las características de éstos, que constituyen un conjunto de datos muy peculiar, para estar así en situación de extraer y utilizar la información disponible de manera óptima. Se pueden enunciar algunos postulados fundamentales sobre la creación y características de los diccionarios, que amplían y concretan la definición que se ofrece al principio del apartado [Har83]:

1. *Los diccionarios se ocupan de la descripción y explicación del vocabulario de una lengua o de una variedad de la misma.*

Dentro de este enunciado se incluyen tanto los diccionarios monolingües, en que la descripción del vocabulario se hace en términos de la propia lengua, como los polilingües, en los que, de una u otra forma, se determinan equivalentes léxicos en uno o varios idiomas. También caben los diccionarios terminológicos, centrados en un dominio determinando.

2. *Las unidades básicas que se consideran en los diccionarios son los 'lexemas', entendidos como combinaciones de forma y significado.*

Este concepto se basa en el modelo de 'palabra' propuesto por Ferdinand de Saussure a principios de este siglo, y que considera que cada palabra consta de dos partes o aspectos

relacionados: por un lado su forma o representación (gráfica/fonética), y, por el otro, su contenido semántico o sentido. El problema de contemplar la palabra bajo este punto de vista es que muy a menudo la relación entre forma y significado no es uno-a-uno, sino que un único signo (forma) puede corresponder a varios significados, o al revés, significados similares pueden corresponder a varias formas diferentes. En el primer grupo (palabras homógrafas) tenemos, por ejemplo, "gato" (animal e instrumento) o "tabla" (objeto de madera o estructura de datos en informática). En el segundo grupo (palabras cuasi-sinónimas) tenemos algunas familias como "árbol-planta-arbusto" o "alto-elevado".

En los diccionarios habituales, la entrada es siempre la forma gráfica de la palabra, y cada uno de los diferentes sentidos que se relacionan con la misma se suele indicar de alguna manera (numerándolos, con sangrados especiales, etc.). También a veces se dan indicaciones de otras palabras semánticamente relacionadas, en el sentido anterior, que conducen a otras entradas. Existen, además, diccionarios específicos de sinónimos, más o menos consistentes [Warn82]. Lo que es un rasgo común a todos ellos es que están orientados a seres humanos que han de aportar su propio conocimiento del mundo, un conocimiento pragmático, para determinar cuál es la acepción que corresponde a sus necesidades, o cuál de los sinónimos presentados es el que realmente se ajusta al contexto.

3. *Los diccionarios pueden describir el vocabulario completo de un lenguaje (o dominio) o concentrarse en uno o más de sus aspectos.*

En los puntos anteriores se ha señalado que en los diccionarios se describe el vocabulario concreto de un lenguaje o un subconjunto de éste, apoyándose en las relaciones existentes entre las distintas palabras. Sin embargo, en la mayoría de las ocasiones no se está en condiciones de abarcar todo el vocabulario de un lenguaje en todos sus aspectos. De manera que, habitualmente, los diccionarios son el resultado de una selección premeditada: palabras más frecuentes, vocabulario técnico, histórico, etc. Incluso en los grandes diccionarios monolingües, si se quiere abarcar varios aspectos de las palabras, el trabajo se suele repartir entre distintos especialistas en cada uno de ellos.

Aunque existen muchos tipos de diccionarios (por temas, usuario al que están destinados, etc.), no hay que olvidar que siempre el vocabulario de un lenguaje refleja el conocimiento que posee el hablante del mundo, del ámbito en que se usa el lenguaje o de la propia sociedad en la que éste se halla. Por lo tanto, muchas veces es difícil separar lo que es información lingüística propiamente dicha e información extralingüística, existiendo palabras (por ejemplo "tabla" o "gato", citadas en el punto anterior) cuyo significado sólo puede explicarse o determinarse en función del contexto en el que ocurren.

4. *En los diccionarios se necesita desarrollar un 'metalenguaje' para manejar y presentar la información.*

Cualquiera que sea el tipo de diccionario que se considere, resulta necesario establecer un marco que modele las relaciones entre las palabras, que no funcionan jamás como elementos aislados, sino que forman parte del contexto más amplio del discurso.

Así, una palabra se puede describir al menos de dos formas: (a) en relación con las estructuras lingüísticas de un cierto número de niveles, que van desde el menor y más simple al mayor y más complejo (fonemas, morfemas, etc.), y (b) en función de las tres dimensiones o aspectos de todo signo lingüístico: semántica (significado), sintaxis (parte de la oración, relación con los otros signos de la secuencia) y pragmática (en relación con el contexto en el que ocurre).

A partir de estos dos modelos, los lexicógrafos han desarrollado herramientas y metodologías para describir de forma precisa las palabras pero, lamentablemente, no existen criterios unificados al respecto.

5. Por último, todos los diccionarios están motivados por y sometidos a las necesidades léxicas del usuario a cuyo uso están destinados.

Evidentemente, diccionarios destinados a distintos usos han de tener formas distintas, y tomar en consideración un conjunto de rasgos que variará de un caso a otro (p.e., en diccionarios técnicos o en diccionarios destinados al uso escolar).

Todos estos postulados son válidos tanto para los diccionarios clásicos como para los diccionarios electrónicos. Estos últimos pueden además caracterizarse más precisamente teniendo en cuenta algunos rasgos significativos. En la Figura 1 se presenta una de las posibles clasificaciones, propuesta por M. Nagao, y que compendia los tipos de DE que existen actualmente [Nag88].

Figura 1. Clasificación de los Diccionarios Electrónicos	
Rasgo	Tipos
Uso	(a) Uso general (b) Terminológicos (c) Específicos por temas
Lenguajes	(a) Monolingües (b) Bilingües (b) Multilingües
Propiedad de la Información	(a) Léxicos específicos (b) Diccionarios en el sentido usual (c) Tesoros (d) Enciclopedias
Usuario	(a) Hombre (b) Máquina

Los tres primeros rasgos que se citan son válidos para todo tipo de diccionarios (electrónicos o no): un diccionario viene caracterizado tanto por el uso al que se vaya a destinar (que determina el tipo de información que ha de contener), como por el número de lenguajes involucrados (si es más de uno, las características del diccionario varían radicalmente) o por las propiedades específicas del léxico que incluye. Esto, que intuitivamente es válido para los diccionarios clásicos, sigue sendo también válido para los DE.

La cuarta característica es la que va a determinar las principales diferencias entre los DE y los clásicos, ya que los requerimientos dictados por el tipo de usuario varían radicalmente en cada caso. Este rasgo también se ha dado en llamar "orientación" [Isa88b] y, como se indica en la tabla, divide a los DE en dos tipos. Tenemos así los DE orientados a ser usados directamente por seres humanos, frente a los DE creados para ser utilizados por una computadora, siendo éste el que va a recuperar la información e interpretarla de forma adecuada.

Los DE que se consideran creados para ser utilizados por el hombre son aquellos en los que el papel de la computadora es el de mero soporte físico de la información; eso sí, con todas las ventajas de almacenamiento, actualización y acceso que se derivan. Son, habitualmente, diccionarios comerciales de "tipo libro" que se presentan en soporte electromagnético u óptico, y pueden incluir una serie de rasgos que sólo son interpretados por un programa y que no afectan a su lectura en pantalla. Pueden ser códigos de cambio de línea, de variación de color o tipo de letra (simulando los distintos tipos de impresión), etc.

Por otra parte, los diccionarios específicamente creados para ser consultados por una computadora sirven habitualmente de base a un programa o conjunto de programas que aplican el conocimiento extraído de éstos a múltiples usos. Entre ellos destacan:

- verificación ortográfica
- verificación sintáctica
- corrección de estilo
- interfaz con un sistema de consulta en lenguaje natural,
- interfaz con un sistema experto,
- análisis para traducción automática
- generación de lenguaje natural a partir de formas codificadas

En este caso el proceso de consulta es automático: el programa es capaz de extraer e interpretar la información que precisa del DE sin intervención humana. Por lo tanto, para que el proceso anterior sea posible, esta información ha de estar codificada de alguna manera, y es este problema, el de la codificación de la información léxica, el que constituye el principal escollo a la hora de desarrollar DE destinados a una utilización automática, por dos razones principales.

En primer lugar, codificar la información léxica implica transformar una información expresada en lenguaje natural (incluyendo el posible metalenguaje del que antes hablamos) en general poco precisa, en un sistema absolutamente concreto y necesariamente conciso. La consecuencia, lamentablemente inevitable, es la restricción del rango de significado de las palabras (ver la Figura 2 en la página 8).

En segundo lugar, como se citaba en el cuarto punto de los postulados, aunque en la modelización de las características de las palabras se tiende a utilizar rasgos de algún tipo, que 'marquen' las palabras, tipificándolas y relacionándolas entre sí, no existe ninguna pauta para ello, y en cada diccionario, glosario o compendio léxico, los mecanismos utilizados son diferentes y,

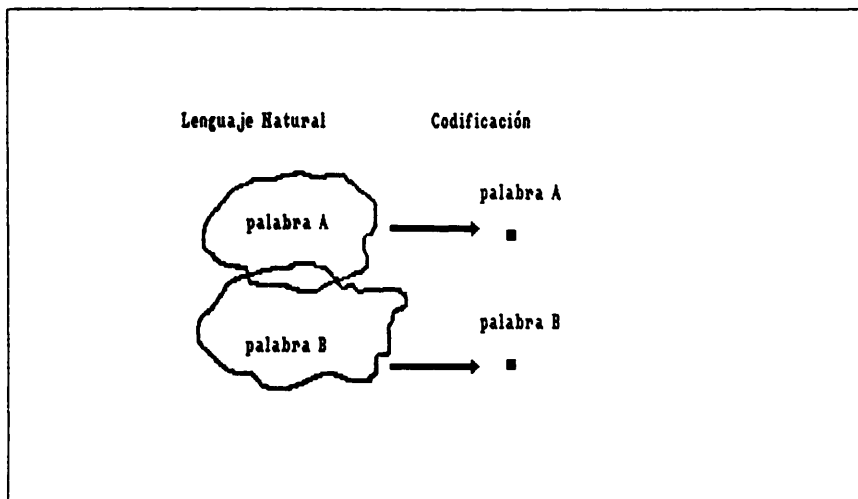


Figura 2. Codificación de la Información Léxica. Restricción del rango semántico de las palabras.

muchas veces, incluso inconsistentes. Para un usuario humano, este problema es fácilmente soluble, porque dispone de un conocimiento propio del mundo que le proporciona un criterio sobre el que apoyarse para resolver las ambigüedades que se le puedan presentar; pero este no es el caso con los programas, por lo que los DE específicamente creados para ser consultados por un programa han de diseñarse teniendo en cuenta y resolviendo todas las posibles ambigüedades que en los diccionarios orientados a los usuarios humanos se dejan a su propia discreción.

Históricamente, ambas orientaciones (a usuarios humanos y programas) se han considerado básicamente distintas e incompatibles, aunque ahora estas diferencias tienden a desenfanzarse, y se comienzan a desarrollar técnicas que permiten utilizar los diccionarios clásicos, destinados a ser usados por el hombre, como punto de partida de ciertas aplicaciones de los sistemas de procesamiento del lenguaje natural.

1.1.2 Codificación Automática

El propósito de los diccionarios, como textos que tratan sobre el lenguaje, es proporcionar definiciones de los sentidos de las palabras y, para cumplir este propósito, no sólo contienen conocimientos puramente lingüísticos, sino que incluyen conocimiento sobre el mundo en general, de tipo semántico y pragmático. Por estas razones, en lingüística computacional, e inteligencia artificial en general, los diccionarios se han contemplado con un gran interés teórico como los instrumentos por

excelencia para investigar la estructura semántica del Lenguaje Natural (LN), y también con un gran interés práctico como los recursos básicos para superar el problema de la adquisición del conocimiento, tanto lingüístico como de cualquier materia específica.

Lamentablemente, en estos momentos la información que proporcionan los diccionarios clásicos no es apta para la mayoría de las aplicaciones que abordan los sistemas de Procesamiento de Lenguaje Natural (PLN); ya que en ellos la información se expresa en lenguaje natural, utilizando una mínima codificación, mientras que los programas necesitan procesar información perfectamente codificada. El problema es, evidentemente, cómo codificar esta información.

Se deduce, por tanto, que los métodos de codificación de la información léxica son actualmente de gran interés, postulándose que ésta puede efectuarse, en general, siguiendo dos aproximaciones diferentes [Wil88]: (a) codificar manualmente una parcela bien acotada del léxico, suficientemente rica, con un propósito determinado, y (b) tratar de desarrollar procedimientos automáticos para transformar el contenido de los diccionarios estándar en una representación adecuada para una gran variedad de aplicaciones del PLN.

En principio, la segunda aproximación parece que presenta muchas ventajas sobre la primera, tanto para las aplicaciones específicas como para la adquisición masiva de conocimiento sobre temas en concreto, pero este proceso no se puede llevar a cabo fácilmente, y también presenta algunos inconvenientes.

La creación de procedimientos automáticos para extraer la información de los diccionarios tiene algunos requerimientos que no necesita la extracción manual, aunque otros son compartidos y afectan directamente al problema de la codificación, se realice ésta como se realice.

Entre las características diferenciadoras de uno y otro tipo de codificación tenemos las siguientes (extraídas en parte de la descripción general del entorno de codificación automática expuesto en [Byr87]):

- Inicialmente, para que se pueda realizar la codificación automática de las entradas de un diccionario es necesario que éste esté disponible en línea, accesible por programa, y, aunque cada vez son más los diccionarios clásicos que se distribuyen en soporte electromagnético u óptico, todavía no se puede contar con todos los que se desearía.

La codificación manual permite, en contrapartida, consultar distintas fuentes (distintos diccionarios) que pueden o no estar disponibles en línea.

- Una operación necesaria, previa al proceso de codificación automática, es la normalización de los datos de partida (las entradas del diccionario), ya que hay que garantizar que los textos de entrada a los programas de extracción son 'regulares', en el sentido de que no existen errores de mecanografiado que pudieran impedir la correcta comprensión de los mismos. Este tipo de errores, que no tienen mayor trascendencia cuando el usuario es humano, pueden conducir a

equivoco a los programas encargados de la lectura automática y, desgraciadamente, no son infrecuentes.

- Para efectuar la codificación automática de la información es necesario analizar el texto de cada entrada; y para ello se ha de utilizar un analizador adecuado. Pero la gramática de las entradas de un diccionario no es la misma que la de las oraciones habituales de un discurso en LN, diferenciándose en tres puntos principales:

1. Las entradas pueden ser grandes porciones de texto (en muchos diccionarios se encuentran entradas que ocupan varias páginas)
2. La definición de lo que se considera una unidad mínima de información varía radicalmente. En los textos corrientes suele ser la oración, mientras que, al considerarse las entradas de un diccionario, la unidad mínima podría ser cada una de las acepciones.
3. La gramática que analiza un diccionario puede ser altamente determinista, dado que el léxico que se utiliza y las estructuras son bastante estereotipadas.

Del primer punto resulta que es necesaria una gran cantidad de espacio en máquina para poder analizar algunas entradas. La consecuencia del segundo es que, mientras en el texto ordinario la delimitación habitual de los elementos de información son los signos de puntuación (puntos, comas, etc.) y los caracteres blancos entre palabras, en las entradas de diccionario hay información asociada a cambios en el tipo de letra (cursivas, negritas), que no necesariamente se presentan con caracteres separadores de ningún tipo. También, aunque estos delimitadores suelen ser identificables, pueden surgir otros problemas dado que a veces hay signos particulares que aparecen entre las palabras (por ejemplo, caracteres especiales para denotar a palabra de la entrada en los ejemplos de uso de la misma, y así ahorrar espacio). El tercer punto sugiere el desarrollo de analizadores no muy complejos, aunque siempre se requerirá el estudio de las estructuras de información utilizadas por cada diccionario en particular.

En cualquier caso, para el desarrollo de métodos, manuales o automáticos, de extracción de la información léxica de las entradas de un diccionario, el más serio obstáculo que se presenta es la codificación de la información semántica que aparece, explícita o implícitamente, en el texto; problema mucho menor en cuanto a la sintaxis se refiere. Esto es debido, en parte, a que la comprensión de la teoría semántica está mucho menos avanzada que la de la teoría sintáctica, como demuestra el hecho de que no exista consenso sobre la mayoría de los principios subyacentes [Hirs89].

En la práctica, esta situación da lugar a una serie de problemas reales que pueden ser de gran envergadura a la hora de extraer información de los diccionarios. Vamos a citar algunos, en los que los datos y marcas a los que se hace referencia se han extraído del diccionario Collins Inglés-Español [Col84]:

- Cada diccionario elige y utiliza marcas diferentes para señalar los distintos usos de cada acepción. Son comunes las marcas de (*fig.*) o (*fam.*) para *sentido figurado* y *acepción coloquial* respec-

tivamente, pero también aparecen otras de carácter morfológico (p.e. *m* o *f* para indicar género), o características más particulares (p.e. *person* para palabras asociadas o aplicadas a personas).

- El conjunto de marcas utilizado no suele ser conocido a priori, por lo que previamente ha de ser detectado a base de filtrar de alguna manera la información contenida en las propias entradas. Por 'marcas' nos estamos refiriendo no sólo a las indicaciones, generalmente entre paréntesis, con las que se caracterizan las entradas (como los ejemplos del punto anterior), sino a todo el conjunto de "*palabras clave*" contenidas en el propio texto de las entradas y que forman, junto con las anteriores, una jerarquía completa de marcas semánticas.
- Incluso dentro de un mismo diccionario, después de determinar el conjunto de marcas utilizado, quedan todavía otros problemas que resolver de índole lógica, sobre la relación (jerárquica o no) que existe entre las marcas.

Muchas veces existe información implícita que ha de extraerse por medio de reglas de herencia, complementariedad o relación. Por ejemplo, aunque no aparezca el género de una palabra en la que se indique que se refiere a *mother*, ésta deberá catalogarse como femenina, ya que *mother* lo es, teniendo que entrar en funcionamiento un mecanismo de relación como el asociado a redes semánticas [Sow84].

Existen también otros casos algo más complejos. Pensemos por ejemplo en la marca *anim*, animado. El que esta marca no aparezca en una entrada no tiene por qué significar que el objeto de la misma sea inanimado, puede que, simplemente, esta información se haya considerado evidente o irrelevante, y por ello no se haya incluido al pensar que usuarios humanos no la necesitarían. Pero para la computadora nada es evidente, y resulta necesario caracterizar la información explícitamente.

Como se deduce de los puntos anteriores, aunque la codificación automática de la información léxica es un tema candente de investigación [Klav88, Nak88, Cai88, Len88], existen todavía muchos problemas, no sólo de índole práctica, sino también teóricos, que han de concretarse antes de que existan resultados realmente útiles en este sentido.

De todo lo anterior se concluye que el principal problema de la extracción de información de los diccionarios, pensando en los DE orientados a los programas, es la codificación de la información, que resulta una tarea muy complicada puesto que no existe ninguna metodología (principalmente semántica) explícita en las propias definiciones. Lo que se utiliza es un conjunto de marcas y signos que son comprensibles por el hombre, ya que éste posee un conocimiento pragmático del mundo. Esta característica de la información léxica hace que la codificación automática de dicha información sea una tarea muy compleja, ya que, además de tener que desarrollar todo un conjunto de programas de normalización, etc., es necesario que los procedimientos de extracción se apoyen en un conjunto sólido de marcas y características basado en las utilizadas en el texto, cuyo establecimiento no es inmediato.

Por lo tanto, para aplicaciones concretas que utilicen un ámbito léxico bien acotado, se piensa que actualmente no es erróneo proceder a la codificación manual de la información, aunque dejando una puerta abierta a futuras innovaciones; la codificación ha de expresarse mediante un formalismo en el que, posteriormente, se pueda transformar la información obtenida mediante procedimientos automáticos.

Éste último es el camino que hemos elegido en nuestro trabajo, en el que se propone un formalismo para la representación de la información bilingüe que admite la codificación manual de las entradas de un diccionario en las primeras fases del desarrollo del proyecto, pero que está diseñado para que se continúe, en fases posteriores, con la explotación automática de los recursos léxicos bilingües disponibles en máquina.

1.2 Traducción Automática y Diccionarios

Bajo el nombre genérico de Traducción Automática (*Machine Translation*) se engloban toda una serie de programas cuya finalidad última es efectivamente la traducción, pero que no siempre realizan este proceso de forma automática [Sio85].

Los sistemas de traducción computerizados se pueden dividir en tres grandes grupos, dependiendo de cuán ambiciosos sean sus objetivos:

- Traducción Automática (TA) propiamente dicha.
- Traducción Asistida por Computadora (TAC).
- Bases de Datos Terminológicas.

En los sistemas de TA puros, la computadora es la única responsable del proceso de traducción: desde el momento en que se le presenta un texto en el lenguaje origen hasta la generación del texto final en el lenguaje meta. Este proceso incluye la gestión completa de las reglas estructurales de transformación así como de todos los diccionarios que pudieran estar involucrados en el proceso (acceso e interpretación de la información), sin que exista intervención humana en ningún momento de la traducción. Este planteamiento no rechaza, sin embargo, la posible pre-edición o post-edición de los textos que se van a traducir, entendiendo por pre-edición el conjunto de tratamientos previos a la traducción (acortamiento de frases excesivamente largas, revisión de la terminología utilizada, etc.) que se pueda hacer sobre los textos, y por post-edición, la fase de corrección (de estilo, etc.) que, en realidad, también incluyen las traducciones humanas. Estos sistemas son, evidentemente, los más ambiciosos.

A continuación tenemos los sistemas de traducción asistida, que a su vez pueden ser de dos clases: sistemas de traducción automática asistida por el hombre (TAAH), *human-assisted machine translation (HAMT)*, y los sistemas de traducción humana asistida por computadora (THAC), *machine-assisted human translation (MAHT)*. Estos sistemas son menos ambiciosos que los anteriores, y el segundo tipo aún menos que el primero. La TAAH incluye aquellos sistemas en que la computadora es la responsable principal de la traducción, pero que, sin embargo, puede interaccionar con un supervisor humano en algunos momentos de la traducción; por ejemplo, para elegir la traducción correcta de alguna palabra en concreto, o determinar términos de coordinación, resolución de referentes, la relación adecuada entre oraciones subordinadas y principal, etc. Frente a este tipo de traducción asistida, en la que la computadora lleva el mayor peso, tenemos la THAC, en la que un traductor humano es el encargado de realizar la traducción por sí mismo, eso sí, en línea con la máquina, utilizando herramientas de tratamiento de textos; de manera que, en ciertas circunstancias, puede requerir los servicios de ésta para acceder a algún tipo de diccionarios especializados, tesauros o bases de datos, o bien para recuperar fragmentos ya traducidos con anterioridad.

Por último tenemos las Bases de Datos Terminológicas, cuyo propósito es mantener un léxico técnico coherente y al día sobre un tema específico. El traductor, que puede o no trabajar en línea, acudirá a estas bases de datos para utilizar la terminología estándar en cada materia. Ésta es una de las principales ventajas del mantenimiento de este tipo de bases de datos: se puede imponer un léxico común en cada área. Pero la estandarización del léxico no es en absoluto trivial, ya que, aunque en ciertos ámbitos se utilice un vocabulario común (dentro de una empresa o grupo de investigación), para que se llegue a una cierta reglamentación, estos grupos han de ponerse de acuerdo entre sí, lo cual no resulta inmediato [Van83].

Cada uno de los tipos de sistemas de traducción mencionados puede constar de varios módulos, cuyas características no son siempre las mismas; lo único que se puede asegurar de todos ellos, de cualquier sistema automatizado de traducción, es que necesitará uno o varios diccionarios que permitan abordar las transformaciones léxicas (elección de términos en la lengua meta) del proceso. De hecho, las bases de datos terminológicas no son más que un tipo específico de diccionario bilingüe, con unos contenidos muy determinados. Estas bases de datos pueden, en principio, incorporarse a los sistemas de traducción asistida, tanto a los TAAH como a los THAC, que, en cualquier caso, necesitan ineludiblemente algún tipo de pauta terminológica.

Sin embargo, no se debe extrapolar que exista compatibilidad "hacia arriba" entre los distintos tipos de sistemas de traducción; es decir, las herramientas diseñadas específicamente para un tipo de sistemas no son en principio aprovechables en los demás. Por ejemplo, en los sistemas de traducción asistida no es siempre necesario disponer de diccionarios orientados a programas, sino de diccionarios orientados a usuarios humanos, ya que, en los casos conflictivos van a ser éstos los que decidan el término adecuado, en función de la información (de tipo textual) que se les proporcione en línea. Por otra parte, en los sistemas de traducción completamente automática la situación es inversa, ya que son los programas los que tienen que acceder e interpretar la información de los diccionarios, por lo que ésta ha de estar adecuadamente codificada.

Otro tema de interés es el conjunto de características de los diccionarios que se requieren en los sistemas de traducción. Éstas varían notablemente de unos tipos a otros, en relación directa con el grado de automatización del proceso, es decir, cuanto más automático sea éste, los diccionarios que se utilicen necesitarán ser más ricos y responder de un mayor número de rasgos lingüísticos, porque los propios programas deberán tomar todas las decisiones relativas al proceso. Por lo tanto, en los sistemas de traducción completamente automática, los diccionarios que se utilicen deben abarcar tanto las transformaciones léxicas entre las palabras del par de lenguas involucrado, como la información morfológica necesaria para el análisis de las palabras de la lengua origen y la síntesis de las de la lengua meta. Vamos a centrarnos en este tipo de sistemas, TA, y profundizar en las características que han de tener sus diccionarios, contemplados como elementos léxicos del sistema.

Conceptualmente, en un sistema de TA es necesario abordar tres tareas, que han de realizarse automáticamente. Estas son: (1) análisis del texto fuente, (2) traducción propiamente dicha, en la que

se encuentran los equivalentes de las palabras y estructuras entre las dos lenguas, y (3) generación del texto final en la lengua meta.

En cada una de estas fases es necesario disponer de un tipo particular de información léxica, cuyas características varían de unas a otras.

En la primera fase, cuando se analiza el texto fuente, el procedimiento léxico que se utiliza ha de ser capaz de determinar, a partir de las palabras flexionadas morfológicamente con prefijos y sufijos, cuáles son sus raíces o bases, para así poder unificar toda la información del resto de las fases en torno a un único elemento, ya que, aunque se ha hecho así en algunos sistemas [Vas85], no es muy ventajoso mantener entradas diferentes en la fase bilingüe para distintas personas o tiempos de un mismo verbo, porque esto implica multiplicar innecesariamente la misma información. Por lo tanto, lo razonable es relacionar cada palabra de la oración original con su forma *base*, y luego ésta con su(s) posibles traducciones.

En el caso del inglés como lengua fuente, que es el utilizado en nuestro trabajo, dada una oración como la del ejemplo (1-1), esta fase léxica inicial de la que estamos hablando debe ser capaz de determinar las bases de las palabras que intervienen, proporcionando la información consecuente:

(1-1) I looked at three flies that were on the table.



BASE: I look at three fly that be on the table

Este proceso puede resultar muy complejo, porque no siempre se puede determinar la base de las palabras aisladamente. Hay ocasiones, como en (1-2), en las que sólo se puede determinar la base de una palabra considerando el contexto. Aquí, la forma "saw" puede corresponder tanto al pasado del verbo inglés "see" ("ver" en español) como es el caso, como al presente del verbo inglés "saw" ("serrar" en español)

(1-2) I saw three flies that were on the table



BASE: I Isee/saw? three fly that be on the table

El conocimiento morfológico de la lengua meta es necesario para poder generar oraciones flexionadas correctamente, en las que se mantengan las características de concordancia propias de

la lengua. Por ejemplo, en español dado un tiempo verbal, una persona y un número ha de poderse determinar la forma verbal correcta:

(1-3) Base: VOLAR	→	"vuela"
Tiempo: Presente		
Modo: Indicativo		
Persona: Tercera		
Número: Singular		

Además de la información anterior, que puede considerarse básicamente de tipo morfológico, es imprescindible contar con diccionarios bilingües que permitan realizar la transferencia léxica entre las palabras y expresiones de uno y otro idioma. Este tipo de información, la información bilingüe, es, por su propia naturaleza, mucho más complejo que el tipo de información que se requiere para el análisis y la generación de los textos origen y meta, ya que obliga a precisar el sentido de cada término, tanto de la lengua origen como de la lengua meta, así como la relación entre ambos. Cómo implementar estos diccionarios, cuál debe ser su estructura y qué tipo de conocimientos deben contener, se convierten en muchos casos en los problemas clave para los sistemas de traducción automática, hasta el punto de que se puede crear o justificar un sistema de TA en función de los diccionarios bilingües que se utilicen, como se ha dado en el proyecto estatal japonés sobre diseño de los diccionarios multilingües [EDR88].

El problema del diccionario bilingüe utilizado resulta entonces de la mayor importancia y, como se demuestra en los trabajos de los distintos grupos de investigación, es un problema que aún no tiene una solución evidente, no ya en cuanto al formalismo de expresión utilizado, sino incluso a otras cuestiones de índole más general [Isa88a]:

1. ¿Se deben utilizar diccionarios de propósito general, o es mejor desarrollar diccionarios dependientes de la aplicación o de un sublenguaje específico?
2. ¿Qué se debe usar como la principal fuente de datos: textos generales sobre cada tema o diccionarios clásicos?
3. ¿La información léxica debe codificarse manualmente o extraerse automáticamente de los textos disponibles?
4. ¿Pueden los diccionarios formularse de forma 'neutral', en el sentido de que pueden ser útiles independientemente de las preferencias teóricas de cada cual?

La respuesta a estas preguntas depende en general del grupo de investigación. Actualmente, por razones prácticas, muchos de los sistemas de TA utilizan un léxico dependiente de la aplicación para la que están diseñados, al menos parcialmente. De esta manera se pueden precisar algunos términos puramente dependientes del dominio, y restringir en alguna medida el problema de la ambigüedad semántica. Por ejemplo, si se está diseñando un sistema para la traducción de textos técnicos informáticos y se trata la palabra inglesa "file", parece que no tiene mucho sentido consi-

derar las acepciones de la misma que no aparecerán en este tipo de textos -"líma", "líla", etc.- y que se puede traducir directamente por "archivo".

Las preguntas 2 y 3 están relacionadas, ya que, como veíamos en el apartado anterior, si los datos se introducen manualmente, entonces es posible consultar varias fuentes simultáneamente, y esto es mucho más difícil en el caso de la codificación automática. Probablemente, la necesidad de extraer las traducciones de unas u otras fuentes dependerá en gran medida del término, ya que hay que tener presente que, sobre todo en ámbitos técnicos, se crean continuamente nuevas acepciones de las palabras, que se recogen en textos relacionados con el tema y no aparecen en los diccionarios terminológicos hasta un tiempo después, o incluso nunca llegan a recogerse en éstos.

En cuanto a la posibilidad de que los diccionarios puedan formularse independientemente de las preferencias teóricas de cada cual, éste resulta ser un problema compartido por casi todos los sistemas de PLN, ya que no existen teorías aceptadas unánimemente y siempre aparecen discrepancias entre los datos que unos y otros grupos consideran. En este sentido se está desarrollando un notable esfuerzo por parte de algunos grupos de investigación hacia la definición de diccionarios *politéóricos* [Wal87, Cal89, Wal90]. La definición de este tipo de diccionarios se apoya en la idea de que, aunque la forma de expresar la información léxica puede parecer estrechamente ligada a una teoría lingüística en concreto, en realidad todos los marcos lingüísticos identifican y describen los mismos hechos léxicos y el mismo tipo de propiedades. Los diccionarios así definidos pueden ser reutilizados para varios fines y a través de diferentes proyectos, evitando una multiplicación innecesaria del mismo trabajo.

En resumen, lo que es evidente es que los sistemas de TA necesitan tres tipos de información léxica diferente: de las lenguas fuente y meta, y bilingüe. Esta situación parece llevarnos de forma natural a la necesidad de tres clases de diccionarios conceptualmente distintos: un diccionario de análisis, uno de transferencia bilingüe (o de traducción) y un tercero necesario para la generación. Esta diferencia conceptual no implica necesariamente que cada una de estas funciones se lleve a cabo por separado, utilizando distintos diccionarios en cada caso, aunque ésta es probablemente la mejor solución desde el punto de vista computacional, ya que resulta más modular y sencilla de mantener.

Actualmente existen algunos sistemas de traducción automática que utilizan un único macro-diccionario al que se accede para determinar las tres acciones descritas [McC88], pero es normal utilizar una estructura de diccionarios que contenga por separado la información monolingüe de los lenguajes origen y meta y la bilingüe que los relaciona, en tres diccionarios diferentes.

1.2.1 Modelo de Transferencia

El modelo de transferencia es el más utilizado actualmente en los sistemas de TA. Este modelo resalta la modularidad del proceso de traducción, potenciando el uso de elementos independientes, que pudieran ser integrados en otros procesos que involucren otros pares de lenguas e incluso en otras aplicaciones relacionadas. Es decir, cada una de las fases del proceso de traducción (análisis del texto fuente, transferencia y generación del texto meta) pueden utilizar elementos independientes entre sí.

En la Figura 3 en la página 19 se presenta un esquema general del proceso de traducción según el modelo de transferencia, indicándose también cuáles son las operaciones que se han de realizar en cada una de las fases propuestas.

Después de la fase de análisis se dispone de una representación intermedia, que es la entrada a la fase de transferencia propiamente dicha. De esta fase se obtiene como salida una representación ya transformada, a partir de la cual se realizará la generación de la oración traducida. Por lo tanto, la fase de transferencia tiene una entidad propia y hace uso de información bilingüe para realizar la transformación de la oración. Algunos sistemas pretenden eliminar incluso la fase intermedia (diferente al análisis y la generación), utilizando para representar la oración formalismos que sean independientes de cualquier lengua natural: la oración origen se analizaría hasta conseguir una representación puramente formal, a partir de la cual se generaría directamente el texto de salida. Se trata de la llamada *interlingua*, que refleja la noción de los "universales lingüísticos". Este modelo presenta enormes ventajas, ya que sólo sería necesario crear, para cada lengua, un módulo analizador y uno generador, y la traducción automática se podría llevar a cabo directamente entre cualquier par de lenguajes, sin necesidad de crear módulos bilingües de transferencia. Por ejemplo, los sistemas *Atlas I* y *Atlas II* de la casa japonesa Fujitsu siguen este planteamiento [Shi89, Uch89]. Sin embargo, el principal problema que presenta el modelo de interlingua es la necesidad de poseer un formalismo completo para la representación del conocimiento lingüístico, válido para todas las lenguas, por lo que suele resultar más asequible el modelo de transferencia intermedia.

En este modelo, aunque toda la información léxica de las dos lenguas pudiera estar contenida en un único diccionario, funcionalmente es más adecuado que existan tres diccionarios distintos:

- un diccionario para el análisis (asociado con un analizador sintáctico-semántico)
- un diccionario para la generación de la lengua meta (asociado a un módulo generador de texto final)
- un diccionario bilingüe de transferencia

Es relativamente sencillo pensar en diccionarios monolingües bidireccionales, es decir, que sean capaces de afrontar tanto los problemas del análisis como de la generación para una única lengua. Lo que ya no resulta tan sencillo es pensar en diccionarios bilingües que sean asimismo bidireccionales, de manera que puedan ser utilizados tanto para la traducción en un sentido como en

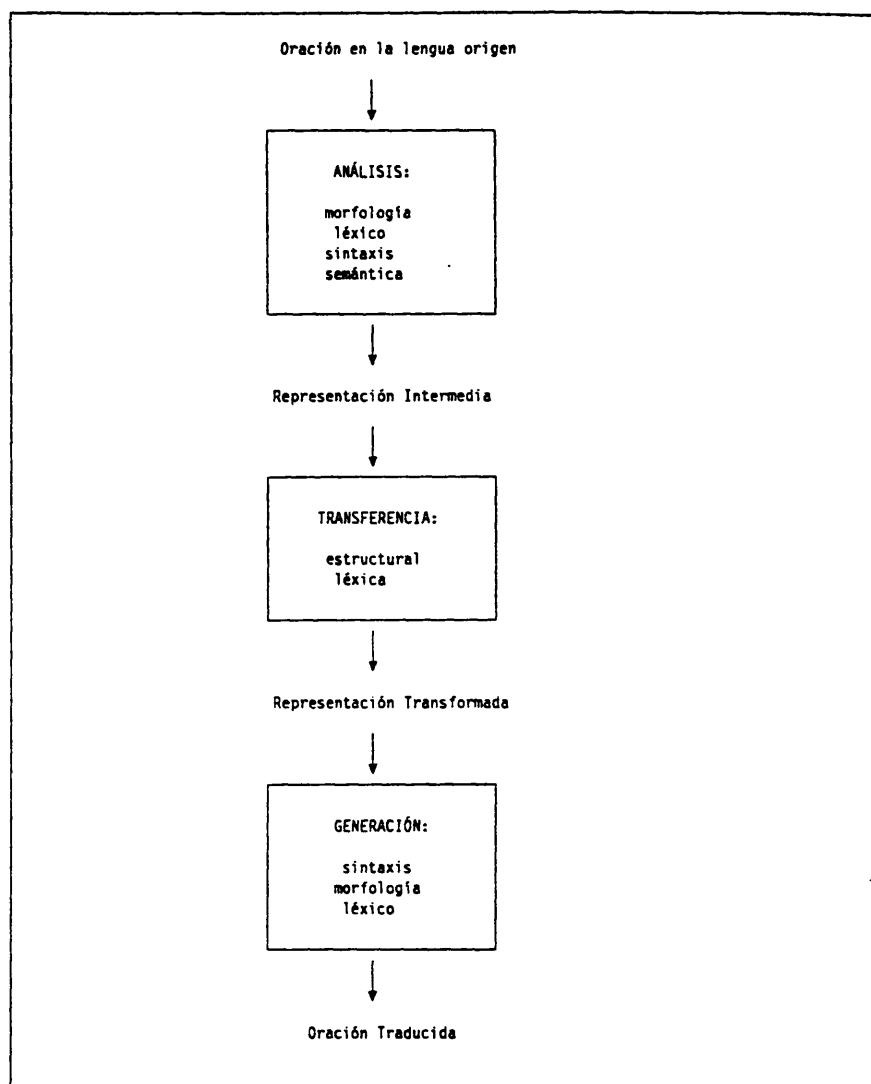


Figura 3. Modelo de transferencia. Esquema del proceso de traducción

el opuesto. La creación de diccionarios de estas características se contempla en el proyecto de traducción automática estatal japonés *MU II* [Tsuj88], refiriéndose a ellos como "diccionarios neutros" o "bidireccionales", nombre que refleja la no direccionalidad de las equivalencias léxicas que contienen. En estos diccionarios se mantendrían las dos correspondencias suprayectivas posibles: de la lengua considerada origen a la lengua meta y viceversa. Pero esta consideración puede ser en cierto modo artificial, ya que no ha de consistir en dos diccionarios diferentes, con una interfaz común, sino en un único diccionario en el que se representen los conceptos semánticos de forma única. Este problema es de gran envergadura, ya que los rangos semánticos de las palabras en distintos lenguajes pueden no coincidir, como veremos más claramente en el ejemplo presentado en la página 60.

En la mayoría de los sistemas, de lo que se dispone es de un diccionario de transferencia direccional, en el que es necesario considerar cuál es la lengua origen y cuál la meta, y que, en principio, no es válido si se invirtiese el sentido de la traducción.

En la figura anterior, que representa el modelo de transferencia, podemos ver también que la caja correspondiente a la fase de transferencia propiamente dicha contempla dos tipos de operaciones: la *Transferencia Estructural* (TE) y la *Transferencia Léxica* (TL). Por *transferencia estructural* se entiende la que gestiona las transformaciones que se producen durante el proceso de traducción en la estructura formal de la oración, y que no están determinadas por la elección de una palabra o expresión concreta como traducción. Un ejemplo muy claro es la transferencia de las oraciones completivas de inglés a español. En inglés sucede que en este tipo de oraciones el nexo, "that", puede elidirse, mientras que en español no sucede así. Por lo tanto todos los nexos de este tipo, en caso de no aparecer, deberán crearse para la oración castellana.

(1-4) I know that he will come.
(1-4t) (Yo) sé que (él) vendrá.

(1-5) I know he will come.
(1-5t) (Yo) sé que (él) vendrá.

En la fase de *transferencia léxica* se incluyen toda la serie de operaciones necesarias para que, dada una palabra o expresión en la lengua origen, se pueda determinar su equivalencia exacta en la lengua meta. La transferencia léxica hace uso, por supuesto, de uno o varios diccionarios bilingües, pero sus características varían considerablemente de unos sistemas a otros. En algunos, esta fase se limita a acceder al diccionario con el único fin de recopilar datos sobre las posibles traducciones, sin decantarse por ninguna de ellas en un principio. En otros, el léxico juega un papel más activo, estando dotado de los mecanismos adecuados para poder seleccionar una traducción concreta en cada caso.

Ambos aspectos de la transferencia (léxica y estructural) pueden o no mantenerse independientes pero, desde nuestro punto de vista, parece evidente que se pueden relacionar muy estre-

chamente. Si se profundiza más en el tipo de operaciones relacionadas con la transferencia, descubrimos que, además de la pura transferencia estructural y de la pura transferencia léxica (en cuanto elección del término o términos equivalentes), existe otro conjunto de operaciones de carácter estructural, pero que están dirigidas por el léxico. Por ejemplo:

(1-6) Mary likes flowers.
(1-6t) A María le gustan las flores.

En esta oración, la transferencia obliga a un cambio en la función de los constituyentes de la oración: el sujeto pasa a ser objeto indirecto y el objeto pasa a ser sujeto. Esta transferencia, que tiene implicaciones estructurales, no es general, sino que tiene lugar únicamente como efecto de la transferencia léxica del verbo principal. Es, por lo tanto, una transferencia estructural dirigida por el léxico.

Este caso, que es bastante frecuente, nos hace suponer una estrecha relación entre ambos tipos de transferencias; siendo muchas veces el léxico el responsable de que tengan lugar o no ciertos cambios estructurales. Ésta es la razón por la que el léxico, a través de los diccionarios bilingües, se convierta a menudo en el núcleo principal de los sistemas de traducción [Pet80, Har83, Ward88].

En el siguiente apartado comentaremos las características de los diccionarios que se vienen utilizando en algunos de los principales sistemas de traducción actuales en fase de investigación o desarrollo, haciendo hincapié en lo que podrían ser los rasgos generales de un diccionario bilingüe para traducción: qué es esencial, qué es deseable y cuáles son los problemas y ventajas de cada uno de ellos.

1.2.2 Transferencia Léxica: Algunos Sistemas

En este apartado vamos a revisar la fase de transferencia léxica de algunos sistemas de traducción automática en fase de comercialización o de desarrollo, principalmente en lo que se refiere a los diccionarios bilingües.

Entre los sistemas que están comercializados actualmente, hablaremos de SYSTRAN, LOGOS y METAL, todos ellos pertenecientes al grupo de sistemas de traducción completamente automática, así como de ALPS TSS Transactive, que es un sistema de traducción asistida (TAAH). También se presenta el mecanismo de EUOTRA, que es un sistema de traducción automática todavía en fase de investigación.

Hay que subrayar la dificultad que existe para recopilar información al respecto por motivos comerciales, porque la documentación disponible es ciertamente escasa, especialmente en lo que a los diccionarios se refiere, ya que suelen ser una de las partes más importantes de los sistemas.

1.2.2.1 ALPS TSS Transactive

ALPS (Automated Language Processing Systems Ltd.) proporciona un conjunto de herramientas integradas para ayuda a la traducción; en este sentido no es un sistema de traducción automática propiamente dicho, sino un sistema de traducción asistida [TSS87].

En cuanto a sus diccionarios, éstos están también distribuidos en varios grupos, siendo en esta ocasión una separación física real. ALPS mantiene archivos diferentes para lo que es el léxico de uso general y la terminología propiamente dicha, determinando además distintos niveles de especialización.

Este sistema posee uno de los diccionarios más simples, ya que cada entrada sólo incluye la parte de la oración que le corresponde y su traducción. Al ser un sistema de traducción asistida por el hombre, se espera que sea el usuario el que decida entre todas las alternativas posibles en cada momento, y elija la adecuada. A continuación se pueden ver dos ejemplos de estas entradas.

```
18 Mbyte
<adj
de 10MB <fixcap
```

```
USER
<n, pers
USUARIO
```

El primer ejemplo corresponde a un caso particular de complementación nominal que se da en inglés, dónde el nombre que actúa como complemento puede asimilarse a un adjetivo, identificándose así en la entrada programada. La traducción al español propuesta aparece en la última línea de la entrada, 'de 10MB', añadiéndose la indicación '<fixcap' para indicar que han de conservarse las letras mayúsculas 'MB'. La segunda entrada es simple, y la única característica especial que tiene es que se marca esta palabra con el código 'pers', de persona, por si este dato pudiese ser de utilidad.

Como se puede ver, la estructura de las entradas definidas en ALPS es muy rígida y no incluye un gran conocimiento lingüístico. Evidentemente, esta forma de describir la información léxica es insuficiente para sistemas más ambiciosos.

1.2.2.2 SYSTRAN

En 1976, la Comunidad Económica Europea comenzó el desarrollo del sistema SYSTRAN de traducción automática entre el par de lengua inglés y francés, a partir de un antiguo prototipo que se había creado a mediados de los sesenta.

Este programa, que ha evolucionado mucho desde entonces, se utiliza actualmente en los departamentos de traducción de la CEE para la traducción de textos técnicos, habiéndose incluido en el sistema la mayoría de los lenguajes comunitarios [Pig88].

La información léxica en SYSTRAN en un principio se definía en un único diccionario, que contenía toda la información de las dos lenguas involucradas en el proceso de traducción: monolingüe y bilingüe. De esta forma había tantos diccionarios como pares de lenguas, existiendo información redundante en ellos, ya que se tenía que incluir la información monolingüe de las dos lenguas en todos los diccionarios de traducción en los que intervinieran.

Posteriormente, estos diccionarios se han reprogramado convirtiéndose en una base de datos de grandes dimensiones, donde cada entrada se define una única vez con todos los códigos que le correspondan (para la desambiguación de la lengua origen y la traducción a las lenguas meta consideradas). Y así para todas las entradas existentes de todas las lenguas disponibles. Si se da el caso de que un lenguaje sólo se utiliza como origen o como meta, los campos correspondientes a sus equivalentes léxicos en las otras lenguas, que no se usan, se mantienen vacíos, y pendientes de una codificación posterior.

En cuanto a los contenidos, SYSTRAN combina en el diccionario bilingüe la información que sólo tiene que ver con la propia traducción, junto con información gramatical y semántica tanto de la lengua fuente como de la lengua meta. Además, parte de esta información está expresada de forma de procedimientos, que se formulan explícitamente, y se utilizan en la fase de transferencia.

Las últimas versiones siguen el modelo de transferencia, tendiendo a mantener la información modular, sin mezclar datos que corresponden a fases diferentes del proceso y codificando en el diccionario bilingüe únicamente la información relativa a la transferencia léxica entre el par de lenguas involucrado, intentando además potenciar la declaratividad de las definiciones.

1.2.2.3 LOGOS

El sistema LOGOS, de la casa del mismo nombre, se demostró por primera vez en 1984, para alemán-inglés en los dos sentidos; y se está trabajando en los prototipos para otros pares de lenguas (francés, español) [Whe85].

En LOGOS la información bilingüe se distribuye en dos diccionarios:

- uno de alta frecuencia, fijo, que contiene palabras funcionales: artículos, preposiciones, etc.
- un diccionario principal, que el usuario puede modificar.

Las palabras contenidas en el segundo de estos diccionarios están calificadas con una serie de códigos ('Área temática', 'Compañía', etc.), que en realidad permiten clasificar la base de datos léxica según distintos criterios (por tema, compañía, producto, etc.).

En LOGOS, de forma similar a SYSTRAN, se utiliza una única interfaz para introducir los datos tanto morfológicos como sintácticos y semánticos, de las palabras de la lengua fuente (por medio de un conjunto bastante elaborado de menús), su traducción (una traducción por 'Área temática'), así como la información morfológica adecuada para la palabra elegida en la lengua meta. Este mecanismo se utiliza para nombres, adjetivos y adverbios. La inclusión de nuevos verbos, que resulta generalmente mucho más compleja, permanece bajo la supervisión directa de los lexicógrafos de la casa LOGOS; por lo tanto, los usuarios no pueden introducir nuevos verbos en el sistema, aunque sí pueden adjudicar diferentes traducciones a los ya existentes, si se varía el 'Área temática'.

Lo que se conoce de los diccionarios de LOGOS puede verse que corresponde al diccionario de terminología diseñado para ser adaptado a las necesidades de los usuarios. De la estructura del diccionario que contiene las palabras de uso general no se tienen datos publicados.

1.2.2.4 METAL

El sistema de traducción automática METAL es uno de los más antiguos utilizados actualmente. Se originó en el Centro de Investigación Lingüística de la Universidad de Texas, en 1961, y ha sobrevivido hasta nuestros días con un pequeño paréntesis. En 1980, Siemens AG compró el prototipo y la investigación y el desarrollo han continuado hasta conseguir que este sistema se convierta en un producto comercial.

Su objetivo principal era la traducción de alemán a inglés, pero potenciando con gran acierto la modularidad de todos sus componentes, tanto lingüísticos como computacionales. Este planteamiento ha llevado al sistema al modelo de transferencia, y ha permitido que sea extensible a otros lenguajes. De hecho, actualmente existen varios prototipos en fase de desarrollo.

Su diccionario bilingüe es de tipo declarativo, pero no sólo posee una relación uno-a-uno entre los diferentes sentidos de las palabras en distintos lenguajes, también contiene información acerca de los cambios estructurales provocados por la determinación de las traducciones, que tendrán lugar si están explícitamente detallados en la entrada correspondiente.

En este sistema el diccionario está dividido por parte de la oración (verbos, nombres, preposiciones, etc.), existiendo tantos sub-diccionarios como el número de éstas. Para cada palabra y cada parte de la oración, se incluyen tantas entradas como posibles traducciones diferentes se prevean. Para cada una de ellas se especifica un conjunto de condiciones que han de verificarse para que se

pueda elegir dicha traducción. También se pueden especificar algunas acciones que han de ejecutarse si esa traducción es la elegida.

El esquema general se presenta en el ejemplo siguiente, extraído del diccionario alemán - inglés [Whi87]:

(schreiben (VST GV) 0	:	write (VST GV) 0)
(T	:	T)
(*	:)
(PP-T0-PP auf on	:)
(PP-T0-PP an on	:)
(NP-T0-PP D to	:)
(+)

En este ejemplo se ve la manera en que se puede forzar la traducción de algunos complementos preposicionales que acompañen al verbo, así como la transformación estructural que sobre ellos se efectúa al traducirse éste. La entrada corresponde al verbo alemán 'schreiben', que puede aparecer complementado por varios PP ó NP. En la entrada se indica que si aparece complementado por un PP introducido por las preposiciones 'auf' o 'an', la estructura se conserva, traduciéndose la preposición por la inglesa 'on' en ambos casos. Si la complementación del verbo es realizada por un NP en caso dativo, este complemento pasa a ser un complemento preposicional en inglés introducido por la preposición 'to'.

Se puede decir que el modelo de diccionario de METAL es uno de los más completos y modulares, aunque no existe mucha documentación al respecto.

1.2.2.5 DLT

Este proyecto de traducción automática, financiado por la compañía BSO, comenzó en 1985 con apoyo del Ministerio de Economía de los Países Bajos. El objetivo es construir un prototipo de sistema de traducción automática de inglés a francés, y está previsto que en 1993 se disponga de una versión comercial [Pap86b, Sch86].

DLT (Distributed Language Translation) es un sistema multilingüe interactivo, diseñado para operar sobre redes de computadores, donde cada máquina actúa como un sistema de traducción independiente a un único lenguaje terminal. Los textos se transmiten entre las distintas estaciones en un lenguaje intermedio (IL), que es una versión de esperanto. Por lo tanto, este sistema sigue en cierto modo el modelo de interlingua, siendo ésta asimilada al esperanto que se utiliza.

Las oraciones que se van a traducir son primero analizadas, creándose múltiples representaciones, ya que el analizador no resuelve los problemas de ambigüedad. Estos árboles son transfe-

ridos a IL y, sobre ellos actúa un sistema 'experto en palabras'. Más tarde se generará el lenguaje requerido a partir de la estructura seleccionada.

En este caso, el diccionario de transferencia es simple, adjudicando directamente una o varias palabras de IL (esperanto) a cada una de las palabras originales. Se crean así múltiples posibilidades, generadas a partir de las combinaciones de las traducciones dadas. La traducción correcta, o combinación correcta de traducciones, se elige a posteriori, por medio del sistema experto en semántica, que verifica cuáles son las relaciones de dependencia correctas entre los candidatos.

1.2.2.6 EUROTRA

Los primeros planes acerca de EUROTRA comenzaron a finales de 1977, creándose en 1978 un grupo permanente de trabajo constituido por miembros de las universidades y demás grupos de investigación europeos. A finales de 1982 se concretaron los planes de trabajo y se puso en marcha este programa, con dos objetivos fundamentales [Kin84]:

- construir un prototipo pre-industrial de sistema de traducción automática, que incluyera todos los lenguajes de la Comunidad Europea
- fomentar la investigación en traducción automática y áreas relacionadas dentro de los países de la Comunidad

Casi siete años después, el primer objetivo parece muy difícil de alcanzar, debido a problemas con la fase de transferencia léxica y al modo en que se han concebido los diccionarios bilingües.

En EUROTRA se ha decidido mantener las fases de transferencia léxica y estructural completamente separadas, además de utilizar un formalismo declarativo para las entradas de su diccionario extremadamente sencillo: únicamente equivalencias entre distintos sentidos de las palabras de cada lenguaje [Hac90]. Esta equivalencia ha de mantenerse entre todos los lenguajes participantes en el proyecto (alemán, danés, español, francés, griego, holandés, inglés, italiano y portugués), de manera que, después de la fase de análisis se hayan desambiguado en función de la lengua origen todos los sentidos posibles de la palabra en todas y cada una de las futuras lenguas meta.

Suponiendo que X sea un cierto lenguaje origen, e Y y Z dos posibles lenguajes meta, ejemplos de entradas en sus respectivos diccionarios bilingües (de X a Y, y de X a Z) podrían ser los siguientes:

Del lenguaje X al lenguaje Y:

t1 = {lu = Xpalabra_sentido1} \Rightarrow {lu = Ypalabra_sentidoa}
t2 = {lu = Xpalabra_sentido2} \Rightarrow {lu = Yparabra_sentidob}

Del lenguaje X al lenguaje Z:

t1 = {lu = Xpalabra_sentido1} => {lu = Zpalabra}
t2 = {lu = Xpalabra_sentido2} => {lu = Zpalabra}

Donde "lu" significa "unidad léxica" (*"lexical unit"*).

Como puede verse, en este caso los posibles sentidos que se pueden desambiguar para cada palabra del lenguaje fuente a veces equivalen a palabras diferentes (en el caso de traducir a la lengua Y) y a veces a la misma (en el caso de traducir a la lengua Z), dependiendo del par de lenguajes considerado. Por lo tanto, cada uno de los sentidos en que es desambiguada una palabra del lenguaje origen se ha de acompañar, cuando los diccionarios son cotejados en los diferentes países, con la documentación adecuada sobre el contexto sintáctico-semántico en el que se identifica dicho sentido, con el fin de que el lexicógrafo que examine los datos pueda establecer las equivalencias apropiadas con su propia lengua, que ahora actúa como lengua meta.

No se han publicado muchos detalles al respecto de este diccionario, y no es muy evidente todavía la forma en que EUROTRA va a gestionar muchas de las implicaciones del proceso de transferencia, entre ellas el conjunto de transferencias estructurales dependientes de las transferencias léxicas.

1.2.3 Recapitulación

Una vez revisados someramente algunos de los sistemas de traducción automática que existen actualmente, podemos extraer algunas de sus características fundamentales, fruto de largos años de experiencia. Entre ellas hay que destacar cómo, comenzando en muchos casos a partir de modelos simples de traducción en los que todas las fases se hacían casi simultáneamente (p.e. SYSTRAN), la mayoría de los sistemas han ido evolucionando hacia modelos más modulares, al menos del tipo de transferencia, en los que el mantenimiento resultase más sencillo y no fuera necesario multiplicar la información monolingüe en función del número de lenguas que se incluya en el sistema. Hay algunos (p.e. DLT) que van aún más allá, aproximándose al modelo de interlingua.

En cuanto a los diccionarios, aunque algunos rasgos de diseño son esencialmente comunes, nos encontramos que los métodos y formalismos varían desde los absolutamente declarativos (EUROTRA) hasta los que incluyen procedimientos en las entradas de las palabras (SYSTRAN). Los hay que mantienen diccionarios independientes para la información monolingüe y bilingüe (que es lo más frecuente) hasta los que tienen toda la información centralizada. También tenemos, desde los que no indican nada en absoluto sobre los efectos que la transferencia provoca en la lengua meta (EUROTRA, p.e.), hasta los que admiten alguna indicación al respecto (METAL).

Pero, en resumen, esta breve revisión, que incluye la evolución histórica de algunos sistemas, nos permite concluir que, independientemente de la implementación, lo que parece más aconsejable desde el punto de vista de la mantenibilidad del diccionario, que no es al fin y al cabo más que un conjunto de datos [Pap86a], es definir las entradas declarativamente, separando de lo que son los datos en sí (condiciones para la discriminación de los sentidos, posibles acciones y traducción de la palabra) de los mecanismos, procedimientos reales, mediante los cuales éstos se ejecutan. Es decir, lo que puedan ser programas y las declaraciones que provocan la ejecución de dichos programas en el momento en que es necesario actuar o recoger más información de la representación de la oración que se está traduciendo.

También es común a la mayoría de los sistemas el diseño de diccionarios "divididos", bien por parte de la oración, o bien por frecuencia de uso de las palabras. Esta decisión es muy razonable, ya que, en cualquier caso, cuanto menor sea el tamaño de los ficheros finales sobre los que se realice la búsqueda de las entradas, mayor será la rapidez del proceso. También es bastante frecuente que se adjudiquen ciertos códigos a las entradas, que en algunos casos puedan hacer de filtro, en función de la aplicación, o servir como medida de peso dependiente asimismo de determinados criterios (área de aplicación, p.e.).

Otro tema muy importante para el lexicógrafo que introduce los datos y para el traductor que hace uso de un sistema de traducción (asistida o no), es el de las interfaces. Como en todos los sistemas que tienen relación con usuarios externos, el entorno de trabajo reviste gran importancia, y es una faceta que ha de cuidarse meticulosamente [Mel88a].

Aunque hemos citado algunos de los sistemas más importantes que existen actualmente, por supuesto, la relación no ha sido exhaustiva y no podemos dejar de nombrar a otros grupos que llevan trabajando intensamente en traducción automática de hace años.

Uno de estos grupos es el liderado por Christian Boitet en GETA (Groupe d'Etudes pour la Traduction Automatique), Universidad de Grenoble, que ha llevado a cabo el desarrollo de varios sistemas de traducción tanto experimentales como comerciales (entre ellos ARIADNE [Boit86]) y que constituyen, más que los exponentes de un sistema en concreto, un centro neurálgico en traducción automática. Es de destacar que este grupo ha abordado recientemente la traducción desde el francés a otras lenguas europeas, respondiendo a los requerimientos de un mercado único en Europa [Boi89].

Entre los sistemas de investigación, en fases más experimentales, son notables los de la universidad de Manchester [Woo87] y, en general, un conjunto de sistemas que conceden una importancia tan notable al léxico que llegan a basar la mayor parte del proceso transferencia en éste. Estas investigaciones, que enlazan directamente con los modelos de representación del conocimiento, tienen un fuerte grupo de seguidores en la universidad de Carnegie-Mellon, Center for Machine Translation [Cul87]. Los sistemas desarrollados por este grupo tienden al modelo de interlingua como modalidad para representar el conocimiento lingüístico, y se refieren a sus diccio-

narios como '*léxicos conceptuales*' u '*ontologías*', donde los nombres se asimilan a marcos de representación, los verbos a sucesos y los adverbios y adjetivos se representan como propiedades de los marcos. [Nyb88, Nir89].

Sobre la transferencia léxica en otros sistemas se puede consultar la documentación del *ISED'88*, que se cita en las referencias, donde se presentan bastantes DE comerciales y en fase de desarrollo (podemos citar [Kaj88, Lan88, Mae88a, Mii88] entre los más importantes).

Teniendo en cuenta la experiencia de los sistemas que hemos descrito se ha desarrollado el prototipo de traducción automática multilingüe MENTOR, cuyo método de transferencia léxica describimos en detalle en los próximos capítulos.

Capítulo 2: Marco de la Investigación: el proyecto MENTOR

2.1 Proyecto MENTOR

A principio de 1987 comienza en el Centro de Investigación UAM-IBM el proyecto de traducción automática MENTOR, siglas que corresponden a Multitarget ENGLISH TranslatOR, y que ya ha sido parcialmente descrito en otros lugares [Rod87, Gol88, Red88, Rod89]. El objetivo de este proyecto es la realización de un prototipo de sistema multilingüe de traducción automática, y su nacimiento corresponde al interés creciente que, como ya se ha mencionado, vienen teniendo todas las aplicaciones del PLN y la traducción automática en particular.

En este proyecto, además del Centro de Investigación de Madrid participan otros grupos de investigación localizados dentro y fuera de IBM: el Centro de Investigación de IBM en Haifa, el grupo científico INESC-IBM de Lisboa y el grupo de Lingüística Computacional de la Universidad de Helsinki.

En el Centro de Investigación UAM-IBM de Madrid existe una gran tradición en lingüística computacional, y se viene trabajando de forma continuada desde hace varios años en aplicaciones de PLN, habiéndose realizado estudios importantes en morfología y sintaxis del español. Por su parte, el grupo de Helsinki es mundialmente conocido por sus trabajos en lingüística, principalmente morfología, siendo los creadores de la llamada *'morfología en dos niveles'* [Kos83].

Entre los objetivos del proyecto se encuentra la formación de grupos competentes en los países involucrados, así como la experimentación exhaustiva en los temas de lingüística computacional relacionados con la traducción automática. No se habían previsto ningunas condiciones en cuanto a la operatividad del prototipo resultante de la investigación. Por varias razones, entre ellas la documentación y herramientas disponibles y la posible aplicación de los resultados obtenidos, se eligió el inglés como lengua origen, siendo las lenguas de los países participantes consideradas "lengua meta" en cada caso (español, hebreo, portugués y finés).

En cuanto a la organización, el trabajo se desarrollaría en cada uno de los centros nacionales, guardando el equilibrio entre la propia investigación de cada grupo, independiente en cierto modo del trabajo de los demás, y la utilización de las mismas herramientas para las fases comunes o compartidas, aunque durante todo el proyecto se cuenta con dos polos principales: Haifa y Madrid.

Se tomaron asimismo algunas decisiones realistas sobre el tipo de textos que se iba a manejar, decidiéndose que el área a tratar sería la documentación técnica de temas computacionales, en concreto de manuales de IBM. Esta decisión ha sido tomada también en otros proyectos similares (ver [Tsu88, Mae88b], por ejemplo), puesto que, al restringir el dominio de aplicación se limita en gran medida la posible ambigüedad semántica de algunas palabras cuyo sentido puede variar considerablemente según el entorno del discurso. Por ejemplo, la ya mencionada palabra inglesa *"table"*, en el entorno de la programación hace referencia generalmente a un tipo de estructura de datos conocido como *"tabla"* en castellano, mientras que en otro tipo de

entornos significará "mesa". Es esta clase de ambigüedades la que se puede resolver limitando el ámbito de aplicación para las traducciones. Además, una decisión de este tipo facilita la elaboración de los diccionarios, puesto que generalmente se puede disponer de un léxico terminológico bien definido, creado inicialmente para los traductores humanos.

La limitación del dominio es uno de los factores que contribuyen a la calidad del texto generado, aunque se admite una fase manual de post-edición que subsane posibles errores y mejore la calidad y el estilo del mismo. Esta posibilidad no invalida el proceso de traducción automática y, en cualquier caso, forma también parte de los procesos de traducción manuales. Sin embargo, no se realizaría pre-edición alguna del texto a traducir, ya que, aunque se podrían simplificar las oraciones para conseguir textos más sencillos, el trabajo que esto supondría sería improbable, además de restar interés a los textos, distorsionándolos artificialmente, cuando uno de los principales objetivos de la TA es tratar material real preexistente, como son, en nuestro caso, los manuales ya elaborados de los que partimos.

Después de casi tres años de trabajo, se han alcanzado los objetivos previstos y el resultado total es muy satisfactorio. En Madrid funciona actualmente un prototipo completo de sistema de traducción automática de inglés a español, capaz de realizar todo el proceso de traducción automática partiendo de oraciones en inglés hasta la generación de las oraciones finales en español. La traducción se realiza al nivel de la oración, no del discurso; es decir, no se resuelven las posibles referencias que excedan el ámbito de la frase (referencias en distintas oraciones de un párrafo). Se maneja un léxico restringido al dominio considerado, aunque se ha experimentado ampliamente sobre las operaciones bilingües de transferencia, parte de cuyos resultados se presentan en esta memoria.

En los apartados siguientes se describen algunas características importantes del sistema desde el punto de vista funcional, que permitirán plantear el problema que nos ocupa, para terminar con un ejemplo de la transferencia léxica para una palabra: todo el proceso a partir de la información contenida en un diccionario clásico hasta la expresión de la información léxica en un pseudocódigo en lenguaje natural. Este ejemplo, que es un caso real de extracción de la información que se ha llevado a cabo en nuestro proyecto, nos servirá para comprobar la complejidad del problema planteado, siendo a la vez el punto de referencia para el formalismo que se expone en el siguiente capítulo.

2.2 Estructura del Sistema

El sistema sigue el modelo de transferencia, del que ya se ha hablado en el capítulo anterior (Figura 3 en la página 19). Este modelo, de acuerdo con la clasificación que hace Tucker según la cantidad de conocimientos lingüísticos que se incorporan [Tuc84], corresponde a la llamada segunda generación de sistemas de TA, en cuanto que se basa en una representación intermedia alrededor de la cual se organiza el análisis y la síntesis, manteniéndose en buena medida independiente de los lenguajes fuente y meta; pero también incluye algunos elementos de la tercera, como son el manejo de información semántica. La frontera entre ambas generaciones no es, sin embargo, muy clara, y es el punto en el que se encuentran actualmente los sistemas de TA más eficientes.

Como ya se ha mencionado, el modelo de transferencia divide el proceso de traducción en tres fases:

- análisis del texto fuente
- transferencia
- generación del texto objeto

La primera fase incluye todo el tratamiento del texto fuente que sea independiente del texto meta y, al finalizar, debe proporcionar una estructura que sea posible transferir y en la que aparezca información suficiente sobre la oración de entrada, resultado de los análisis que se hayan efectuado.

En nuestro sistema, para esta fase se utilizan diversas herramientas desarrolladas en el laboratorio T.J. Watson de IBM en Yorktown Heights dentro de otros proyectos [Jen86]: un lenguaje específico para el tratamiento del lenguaje natural y su procesador asociado, PLNLP, una gramática del inglés, PEG, y un amplio léxico del inglés, OD (basado en los diccionarios Webster-7 y Longman).

PLNLP, siglas correspondientes a Programming Language for Natural Language Processing, es un lenguaje especialmente diseñado para el tratamiento declarativo del lenguaje natural [Heid72]. Los módulos escritos en este lenguaje pueden ser posteriormente compilados a distintos lenguajes de programación, como PL/B y LISP/VM. Es en este lenguaje en el que está escrita PEG, siglas de PLNLP English Grammar [Jen87]. El análisis que esta gramática realiza de las oraciones es general, independiente del área de aplicación, y de una gran cobertura sintáctica. Entre sus características más señaladas se encuentra la posibilidad de producir, cuando no es capaz de analizar por completo una oración, lo que llama "análisis aproximados" a base de ajustar distintos fragmentos del texto que constituyan una unidad [Jen83]. De esta forma siempre hay disponible un análisis para cualquier frase propuesta, en el que se indica que es "aproximado"; de forma que es posible tomar decisiones sobre estos análisis sin necesidad de interrumpir el proceso global (intentar traducirlos en cualquier caso, dejar la oración original, dar algún tipo de mensaje, etc.).

Para cada oración, PEG proporciona uno (o varios) grafos orientados en los que se representan los resultados obtenidos tras el análisis. Se genera más de un grafo cuando la oración de entrada presenta algún tipo de ambigüedad, calculándose un coeficiente para cada uno de ellos, basado en determinadas características del análisis, y que indica su plausibilidad [Hei82]. En estos casos de ambigüedad, seleccionamos para el prototipo el análisis más probable según PEG, de manera que a partir de ahora consideraremos que tenemos una entrada única a la siguiente fase de traducción.

Aunque PLNLP es un analizador basado fundamentalmente en la sintaxis, su diseño permite la adición de una o varias fases de procesamiento semántico, adaptadas a las necesidades y características específicas de cada aplicación. Como resultado, el árbol de análisis desambiguado lleva asociadas explícitamente informaciones de la estructura profunda junto con la información superficial que caracteriza a la oración fuente, ambas necesarias para generar correctamente la salida, ya que el resultado que el sistema genere debe reproducir el sentido del original, pero manteniendo también, al mismo tiempo, las características estructurales y estilísticas de aquél hasta donde sea posible.

Esta primera etapa de análisis es general y compartida, cualquiera que sea el lenguaje que se va a generar. Sólo la segunda fase, la de transferencia, empieza a utilizar información del lenguaje terminal. En ella, la estructura de análisis antes obtenida se modifica de modo que refleje la estructura de la lengua meta. Se ha de emplear un diccionario bilingüe que contenga las equivalencias de cada término, con el objeto de construir el esqueleto básico en el que se encuentre la información sintáctica y semántica para la última fase.

El proceso de transferencia [Red89a, Red89b] consta fundamentalmente de las dos partes de las que ya hablamos en el capítulo anterior: transferencia léxica y transferencia estructural. El tamaño y las características de esta fase varían enormemente de unos sistemas a otros, pero siempre incluye acciones de desplazamiento y reordenación de los elementos de la oración, junto con la desaparición de algunos elementos y la aparición de nuevos huecos, en algunos casos.

En nuestro prototipo la fase de transferencia comienza sobre el resultado de los análisis de PEG. Este resultado es un grafo orientado a partir del cuál generamos una estructura arborescente decorada, siendo esta representación la que se mantiene durante todo el proceso de traducción. La principal ventaja de la representación arborescente es su sencillez, lo que la hace muy apta para las fases de investigación y pruebas.

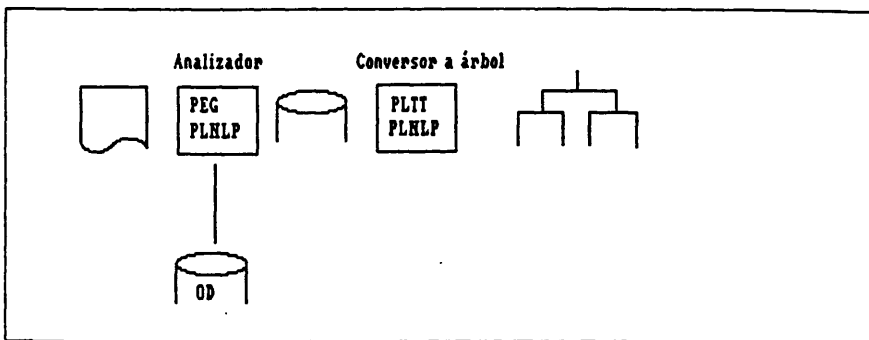


Figura 4. MENTOR: análisis.

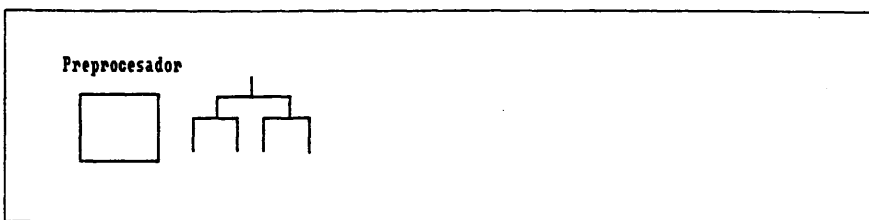


Figura 5. MENTOR: preprocesamiento.

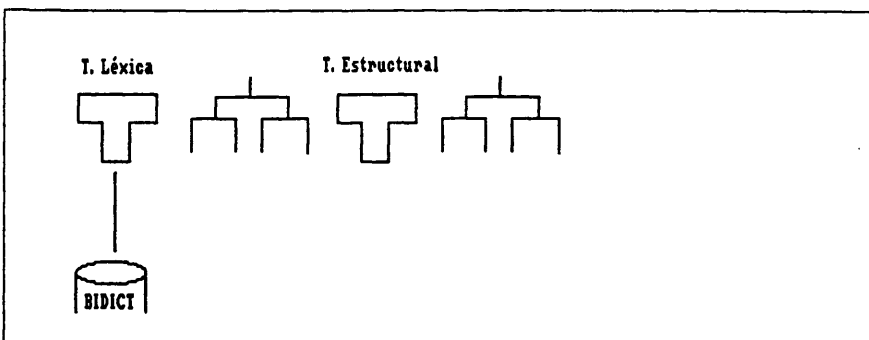


Figura 6. MENTOR: transferencia.

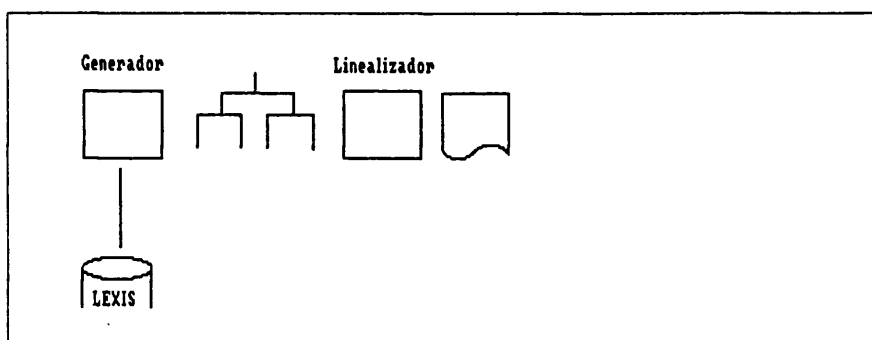


Figura 7. MENTOR: generación.

Después de la transferencia se dispone de otro árbol decorado a partir del cual comienza la fase de generación o síntesis. En esta fase se crea la oración de salida a base de recorrer la estructura previamente construida, adaptando la información a la forma superficial que se utiliza habitualmente en español y flexionando las palabras de la manera adecuada. Para esto se utiliza un diccionario específico que es fruto de los trabajos realizados por nuestro grupo para otro proyecto [Cas85, Cas87], y que se puede considerar completo en lo que a morfología del español se refiere.

El sistema se ha desarrollado en el entorno del sistema operativo VM/CMS, y en cuanto al lenguaje de programación, el núcleo fundamental del prototipo está implementado en LISP/VM; comenzando por el analizador PEG, que se utiliza en la versión compilada a LISP. Para la implementación de las reglas de control y transferencia, se dispone de un paquete de funciones específicas para la transducción de árboles [Nom87] también en LISP. Esta utilidad dispone de funciones para la búsqueda y reconocimiento de patrones en los árboles, así como para la modificación de su estructura, consulta de propiedades, inserción y eliminación de nodos, etc.

Los diccionarios utilizados (para análisis, generación y transferencia) conforman módulos independientes, externos a las reglas y a los propios programas. La comunicación entre éstos y los diccionarios se realiza a través de una serie de programas de interfaz realizados en REXX. En este lenguaje se han implementado asimismo todo el resto de interfaces que ha sido necesario desarrollar para el sistema.

En las figuras anteriores, a las que se hará referencia en los siguientes apartados, se detalla gráficamente el proceso de traducción que se efectúa en MENTOR, desde la oración original hasta la ya traducida. Cada una de ellas corresponde a una fase del proceso: análisis de la lengua origen, preprocesamiento de la información previo a la transferencia, transferencia propiamente dicha y generación. La representación es esquemática, habiéndose dispuesto la entrada y salida del pro-

gramática por medio de ficheros que contienen las oraciones a traducir y las ya traducidas, sin embargo la entrada y, consecuentemente, la salida, también pueden realizarse de forma interactiva.

2.2.1 El Analizador

Como ya se ha mencionado, PLNLP es una de las herramientas fundamentales de nuestro proyecto. Las unidades básicas del lenguaje son las reglas y los registros. Los registros son colecciones de atributos y valores, pudiendo éstos ser también apuntadores a otros registros, creándose así una compleja red de informaciones. Además PLNLP también soporta listas, cadenas, etc. En cuanto a las reglas, corresponden al tipo de estructura de frase aumentada, y pueden ser de dos tipos, según sirvan para "decodificar" (analizar), o "codificar" (generar). Con cada tipo de regla hay asociado un algoritmo diferente: en el primer caso el proceso es de abajo arriba y en paralelo, en el segundo de arriba abajo y en serie. Sin embargo, la estructura básica de las reglas es la misma en ambos casos: hay una parte izquierda en la que se identifican los constituyentes y se especifican las condiciones de activación de la regla, y una parte derecha, separada de la anterior por una flecha, en la que se identifican los nuevos constituyentes y se especifica la nueva estructura resultante de la aplicación de la regla.

PLNLP dispone de un entorno completo que facilita la tarea de desarrollo de una gramática, permite visualizar árboles de análisis, dispone de facilidades de traza, de selección de reglas de la gramática que cumplan ciertos requisitos, etc.

La gramática del inglés, PEG, genera para cada oración una descripción sintáctica en forma de grafo con estructura atributo-valor, y a partir de ella construye y muestra al usuario un árbol de análisis más legible, utilizando una parte de la información de dicha estructura. Para ello PEG dispone de tres conjuntos de reglas:

- reglas del núcleo gramatical, que definen las estructuras centrales y generales del inglés
- reglas de manejo de ambigüedades, que se encargan de decidir, en casos de análisis múltiples, cuál de todos es el preferido
- reglas de recuperación de fallos, que asignan una estructura "razonable" a la entrada en los casos en que las reglas del núcleo no son capaces de analizar la frase (los "análisis aproximados" que mencionamos en el apartado anterior)

```

DECL1 NP1 PRON1* "I"
      VERB1* "allow"
      VP1 COMPL1 "that"
          NP2 PRON2* "I"
          VERB2* "made"
          NP3 DETP1 ADJ1* "a"
              NOUN1* "mistake"
PUNC1 ".*"

```

Figura 8. Árbol de análisis proporcionado por PEG. "I allow that I made a mistake"

```

SEGTYPE 'VERB'
STR      " allow"
COPYOF   REC "allow"
BASE     'ALLOW'
DICT     'allow'
POS      VERB
INDIC    PLUR PRES INF DITRAN INGCOMP NPTOV THATCOMP TRAN
OBJTPREP 'FOR'
OBJTPREP 'OF'
PRED     'ALLOW'
NODENAME 'VERB1'

```

Figura 9. Ejemplo de un registro de la estructura: el verbo. "I allow that I made a mistake"

Al comenzar la fase análisis se crea un registro para cada palabra, que contiene todos los datos que el diccionario posee de la entrada correspondiente. Dicho registro se enriquece progresivamente según se van aplicando las reglas de la gramática, creándose nuevos registros de nivel superior según progresa el algoritmo de análisis. Por ejemplo, sobre un nombre se creará un registro de NP (Noun Phrase), que a su vez puede formar parte de otro constituyente, etc. La aplicación de una determinada regla a un conjunto particular de constituyentes la determina la presencia o ausencia de ciertos atributos y sus valores en los registros. Algunos de estos atributos los proporciona el diccionario, mientras que otros los añaden las propias reglas. Lo que PEG hace, fundamentalmente, es producir una descripción sintáctica de la oración, que se construye progresivamente. Después de procesar una oración, se visualiza el árbol de análisis correspondiente (ver la Figura 8), que se forma con la información básica de cada elemento: parte de la oración o tipo de constituyente y palabra. Este árbol tiene la ventaja de ser conciso y fácilmente legible pero, como contrapartida, contiene mucha menos información que el conjunto completo de registros (ver la Figura 9 y la Figura 10 en la página 42), ya que el árbol que PEG muestra no es en realidad más que un esquema de la estructura. En los registros se encuentran otros datos que, por supuesto, son

necesarios para las siguientes fases del proceso. En MENTOR, a partir de la estructura original se extraen parte de estos datos, formándose un *árbol decorado* como el que puede verse en la Figura 11 en la página 43. A partir de ahora haremos referencia exclusivamente a la representación de la información contenida en el árbol decorado y que es el punto de partida de la fase de transferencia.

Para extraer esta información de los registros se ha escrito un pequeño procedimiento en PLNLP (versión LISP), llamado PLTT, cuya actuación se esquematizaba en la Figura 4 en la página 37, después del proceso de análisis. En dicho procedimiento se pueden indicar fácilmente los datos que se quiere considerar de toda la información disponible. La forma de expresar esta información en el árbol depende del tipo de los datos seleccionados, y se detalla a continuación, correspondiendo todos los ejemplos que se mencionan al árbol extraído para la oración *"I allow that I made a mistake"* representado en la Figura 11 en la página 43.

1. Si los datos seleccionados son simples indicadores, representados por INDIC en los registros de PEG, como puede verse en la Figura 9 en la página 40 y en la Figura 10 en la página 42, entonces se incluyen en el nuevo árbol como un par atributo-valor en el que, o bien el nombre del atributo es el nombre del indicador y el valor es siempre '1' (p.e. (ANIM . 1)), o bien, en aquellos casos en que puedan existir varios valores para el mismo atributo (PERS1, 2 o 3, p.e.), el nombre del atributo es el genérico de la característica (PERSON, p.e.) y el valor el que corresponda.

En el caso concreto de un nodo de la oración propuesta, N1 por ejemplo, que corresponde a la oración en su conjunto, tenemos que la característica transitivo, TRAN, que aparecía como indicador en los registros (ver la Figura 10 en la página 42), se ha convertido ahora en un par (TRAN . 1), mientras que el número, que también aparecía como indicador, al tener más de una posibilidad (singular ó plural) se ha representado por un par (P-NUM . SING), donde P-NUM es el nombre elegido para esta característica.

2. En el caso de campos variables por definición, cuyo valor en los registros es una etiqueta o una constante (como es el caso de la BASE, que depende de cada palabra), si aparecen una única vez la representación es también un par atributo-valor, como en el caso anterior. Si puede haber varios campos con el mismo nombre, entonces es una lista en la que el nombre del campo aparece en primer lugar y los valores posibles a continuación (p.e. OBJPREP).

Para ejemplificar la representación de este tipo de datos nos podemos fijar en el nodo N4, que corresponde al verbo principal de la oración. En este nodo, al igual que en todos los demás ya que la BASE es un dato siempre presente, tenemos un par (BASE . ALLOW) que corresponde al primer caso. El segundo nos lo encontramos en otra característica de este mismo nodo, (OBJPREP OF FOR), que indica las preposiciones que este verbo admite como introductoras del objeto directo. Estas preposiciones estaban indicadas en sendos campos del registro correspondiente, que puede verse en la Figura 9 en la página 40.

3. Por último tenemos el caso de los apuntadores. Al transformar el grafo en un árbol no se pueden mantener más relaciones de carácter apuntador que las puramente jerárquicas, por propia cons-

```

SEGTYPE 'SENT'
SEGTYPE2 'DECL'
STR      " i allow that I made a mistake"
RULES    4000 4140 5000 7200
RULE     7200 REC VP2 PUNC1
COPYOF   VP2 "i allow that I made a mistake"
BASE     'ALLOW'
DICT     'allow'
POS      VERB
INDIC    SING PRES INF INGCOMP NPTOV THATCOMP TRAN PERS1
PRMODS   NP1 "i"
HEAD     VERB1 "allow"
PSMODS   VP1 "that I made a mistake"
PSMODS   PUNC1 "."
FRSTV    VERB1 "allow"
SUBJECT  NP1 "i"
OBJECT   VP1 "that I made a mistake"
PREDCOMP VP1 "that I made a mistake"
NARGS    1
PARSENO  1
SENTE    'DECL'
TOPIC    NP1 "i"
PRE      'ALLOW'
DSUBJECT NP1 "i"
DOBJECT  VP1 "that I made a mistake"
NODENAME 'DECL1'

```

Figura 10. Ejemplo de un registro de la estructura: la oración. "i allow that i made a mistake"

trucción del árbol. Por lo tanto se pierden los apuntadores que en el grafo reflejan otro tipo de relaciones (sujeto de la oración, núcleo del nivel, partícula asociada a un verbo, etc.).

Para poder disponer de la información relativa a estas relaciones, que ha sido determinada durante el análisis, lo que se hace es marcar de cierta forma tanto el nodo inicial como el final del arco.

Tomemos por ejemplo el nodo de la oración en su conjunto, N1, cuyo registro se detalla. Allí podemos ver que existe un campo denominado SUBJECT que apunta a NP1. La existencia de este apuntador supone que en el nodo N1 del árbol, el de la oración, se incluirá un par atributo-valor en el que el nombre del atributo es HASSUBJECT y el del valor NP1: (HASSUBJECT . NP1), y en el nodo correspondiente a NP1, N2 en la estructura del árbol, aparecerá un par simétrico indicando que ese constituyente actúa de sujeto: (ISASUBJECT . 1). El mismo mecanismo se utiliza para el objeto directo, indirecto, etc.

Con la transformación descrita de los registros de PEG al árbol, realizada por el programa PLTT, termina la fase de análisis. A continuación se realiza un sencillo preproceso, anterior a la fase de transferencia, que actúa sobre la estructura de árbol obtenida, Figura 5 en la página 37.

La finalidad de este programa es simplificar algunos de los rasgos que existen en el árbol y proporcionar a la fase de transferencia una representación más compacta de la información.

Por ejemplo, una de las acciones que se llevan a cabo es la fusión de todos los nodos verbales que se crean para tiempos complejos en uno sólo que condensa toda la información. Así, si en una oración se utiliza una forma perfecta como "has made", los dos nodos que se habrán creado, uno correspondiente a "has" y otro correspondiente a "made", se condensan en uno sólo, cuya BASE será

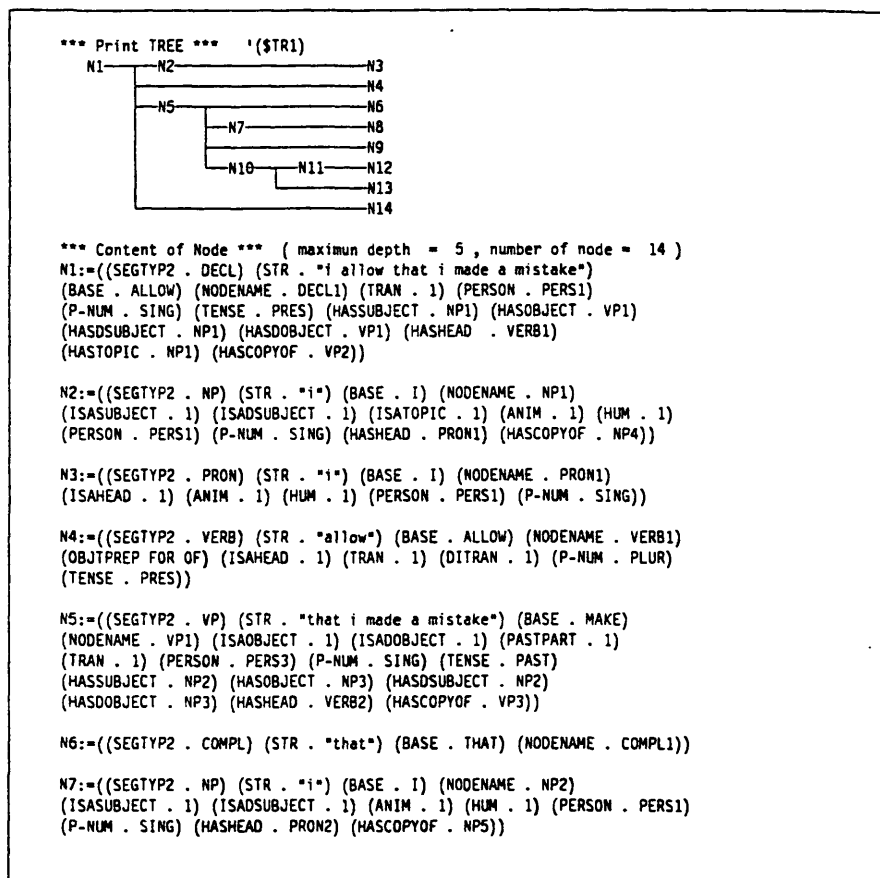


Figura 11. Representación al comienzo de la fase de transferencia. "I allow that I made a mistake"


```

N8:=((SEGTP2 . PRON) (STR . "I") (BASE . I) (NODENAME . PRON2)
(ISAHEAD . 1) (ANIM . 1) (HUM . 1) (PERSON . PERS1) (P-NUM . SING))

N9:=((SEGTP2 . VERB) (STR . "made") (BASE . MAKE) (NODENAME . VERB2)
(OBJTPREP AFTER FOR TOWARD TOWARDS WITH) (PTC OUT UP) (ISAHEAD . 1)
(PASTPART . 1) (TRAN . 1) (DITRAN . 1) (TENSE . PAST))

N10:=((SEGTP2 . NP) (STR . "a mistake") (BASE . MISTAKE)
(NODENAME . NP3) (ISAOBJECT . 1) (ISADOBJECT . 1) (ART . 1)
(PERSON . PERS3) (P-NUM . SING) (NO-CALIF . DET) (DEFIND . INDEF)
(HASHEAD . NOUN1) (HASCOPYOF . NP6))

N11:=((SEGTP2 . DETP) (STR . "a") (BASE . A) (NODENAME . DETP1)
(ART . 1) (P-NUM . SING) (NO-CALIF . DET) (DEFIND . INDEF)
(HASHEAD . ADJ1) (HASCOPYOF . AJP1))

N12:=((SEGTP2 . ADJ) (STR . "a") (BASE . A) (NODENAME . ADJ1)
(ISAHEAD . 1) (ART . 1) (P-NUM . SING) (NO-CALIF . DET) (DEFIND . INDEF))

N13:=((SEGTP2 . NOUN) (STR . "mistake") (BASE . MISTAKE)
(NODENAME . NOUN1) (ISAHEAD . 1) (PERSON . PERS3) (P-NUM . SING))

N14:=((SEGTP2 . PUNC) (STR . ".") (BASE . @.) (NODENAME . PUNC1)
(HASCOPYOF . PUNC2))

```

Figura 12. Representación al comienzo de la fase de transferencia (cont.). "I allow that I made a mistake"

MAKE, que es el verbo principal, y que tendrá como rasgo (PERFECT . 1). Esta información es suficiente para el tratamiento posterior que se realiza en las fases de transferencia y generación, y la representación condensada resulta más eficiente tanto desde el punto de vista computacional como de diseño.

2.2.2 La Transferencia

La transferencia comienza donde finaliza la fase de análisis y termina donde comienza la síntesis de la oración final. Esta afirmación es evidente desde un punto de vista teórico, pero lo que no está nada claro es cuáles son estos dos extremos, es decir, cuáles son las funciones que racionalmente se supone deben contemplar las fases anterior y posterior a la transferencia:

- ¿debe el análisis proporcionar información funcional sobre los constituyentes de la oración? (objetos, p.e.)
- ¿debe proporcionar sólo los principales elementos funcionales, "calculándose" el resto durante la transferencia? (el nombre al que modifica un adjetivo, p.e.)

- ¿la presencia y orden de los premodificadores nominales pertenece a la transferencia, o debe relegarse a la fase de generación? (orden de nombre y adjetivo, p.e.)
- ¿el tratamiento de las preposiciones es generación o transferencia?

Y se pueden seguir enumerando temas que resultan conflictivos, y sobre los que no hay ninguna regla preestablecida. Parece obvio que se debería intentar llevar el análisis lo más lejos posible, y que la síntesis debiera ser capaz de gestionar cuantas más estructuras mejor; sin embargo hay que hacer esta afirmación con cuidado, porque una de las grandes dificultades que existen en el PLN es la de afrontar la implementación práctica de los modelos teóricos. Muchas veces hay que llegar a soluciones de compromiso. De hecho, la experiencia indica que tiene sentido plantear sistemas que mantengan un cierto equilibrio entre lo teóricamente aconsejable y las necesidades prácticas de conseguir modelos que funcionen con unos resultados aceptables y en un tiempo razonable. La solución es fijar criterios internos que permitan saber a qué fase han de pertenecer los distintos tipos de operaciones cuando se tome en consideración su implementación. En nuestro grupo se consideran pertenecientes a la fase de análisis todas las operaciones que PEG lleva actualmente a cabo, asumiendo que sería conveniente incluir aún otras más, por ejemplo, se considera necesario profundizar más en el análisis semántico de los constituyentes, ya que PEG todavía no proporciona suficiente información en este sentido.

Creemos que sería necesario incluir en esta fase el análisis de todas las relaciones profundas entre los distintos elementos de la oración. PEG ya realiza el tratamiento de las relaciones profundas entre elementos funcionales, determinando y diferenciando el papel superficial y profundo que éstos adquieren. Esto permite que en una oración como la siguiente:

(2-1) The girl was given a flower.

"The girl" sea identificado como sujeto superficial, SUBJECT, y objeto indirecto profundo, DOBJECT, mientras que el sujeto profundo, DSUBJECT, queda vacío y "a flower" se identifica como el objeto profundo de la oración. Estos rasgos están señalados en el registro de la oración que se presenta, junto con el árbol, en la Figura 13 en la página 46.

En cuanto a la generación, hemos decidido que sea lo mayor y más completa posible, de forma que realice no sólo la flexión morfológica de las palabras, sino también la adaptación de la estructura superficial de la oración: reordenando sus elementos, eliminando y creando aquellos que sean necesarios, etc. Quedan entonces para la fase de transferencia las operaciones que son *puramente* bilingües:

DECL1	NP1	DETP1	ADJ1*	"the"
		NOUN1*	"girl"	
	AUXP1	VERB1*	"was"	
	VERB2*	"given"		
	NP2	DETP2	ADJ2*	"a"
		NOUN2*	"flower"	
	PUNC1	"."		

SEGTYPE	'SENT'
SEGTYPE2	'DECL'
STR	" the girl was given a flower"
RULES	4000 4050 4010 5000 7200
RULE	7200 REC VP1 PUNC1
COPYOF	VP1 "the girl was given a flower"
BASE	'GIVE'
DICT	'given'
POS	VERB NOUN ADJ
INDIC	SING PAST PASSIVE DITRAN NPTOV TRAN PERS3 CLOSED
PRMODS	NP1 "the girl"
PRMODS	AUXP1 "was"
HEAD	VERB2 "given"
PSMODS	NP2 "a flower"
PSMODS	PUNC1 "."
FRSTV	VERB1 "was"
=> SUBJECT	NP1 "the girl"
=> OBJECT	NP2 "a flower"
NOUN	REC "given"
ADJ	REC "given"
NARGS	3
AUXFORM	1
PARSENO	1
SENTYPE	'DECL'
TOPIC	NP1 "the girl"
PRED	'GIVE'
=> DOBJECT	NP2 "a flower"
=> DINDOBJ	NP1 "the girl"
=> DSUBJECT	REC1 ""
XPASSIV	1
XVPMODS	1
XCLEAN	1
NODENAME	'DECL1'

Figura 13. Ejemplo de detección de funciones profundas en la oración. "The girl was given a flower"

1. elección de las palabras equivalentes en cada caso
2. transformación de las estructural gramaticales típicas del inglés a sus equivalentes en español

La primera de estas operaciones se identifica con la **transferencia léxica** y la segunda con la **transferencia estructural**. Aunque, como ya se mencionó en el capítulo anterior, se puede pensar en mantener ambos tipos de operaciones completamente separadas, nosotros creemos que esto no es aconsejable ya que existen de hecho transformaciones estructurales dirigidas directamente por el léxico. Pensemos, por ejemplo, en las originadas durante la transferencia de ciertas preposiciones:

- (2-2) You will get it by writing a letter to his manager.
- (2-3) Don't worry, it will be finished before going out.

que se traducen:

- (2-2t) (Usted) lo conseguirá escribiendo una carta a su director.
- (2-3t) No se preocupe, estará terminado antes de salir.

En los dos casos anteriores aparecen sendas construcciones de oraciones preposicionales, "*by writting a letter*" y "*before going out*", que, sin embargo dan lugar a estructuras diferentes al realizarse la transferencia al español. En el nivel de profundidad de análisis que nos movemos y desde el punto de vista únicamente del inglés, no hay razón para distinguir los casos (2-1) y (2-2), siendo ésta tarea de la transferencia. Y, aunque el resultado es una transferencia estructural, ésta esta provocada directamente por la ocurrencia léxica de preposiciones distintas. Por lo tanto, parece coherente que sea la ocurrencia de dichas preposiciones la que active las reglas de modificación estructural necesarias, en el momento en que se acceda al nodo correspondiente a la preposición (que podría ser, en los ejemplos, el de toda la oración subordinada o el de la preposición simple). Por otra parte, también resulta sensato plantear el sistema de forma que cada nodo de la estructura sea visitado el menor número de veces posible, es decir, que sólo se realice una "pasada" durante la transferencia.

Teniendo en cuenta los factores anteriores, se ha decidido que la información contenida en el diccionario bilingüe para cada entrada pueda también indicar y activar los cambios estructurales que se originen por el hecho de elegir una cierta traducción.

Frente a este tipo de cambios estructurales dirigidos por el léxico, tenemos otros que son de carácter general, no dependiendo de la ocurrencia de ninguna palabra en concreto. Un claro ejemplo es el tratamiento de las estructuras de genitivo en la transferencia del inglés al español:

- (2-4) Mary's apartment is really very nice.

Este tipo de transformación ha de realizarse siempre, independientemente de quién sea el poseedor y de las características de la cosa poseída.

- (2-4t) El apartamento de Mary es realmente muy bonito.

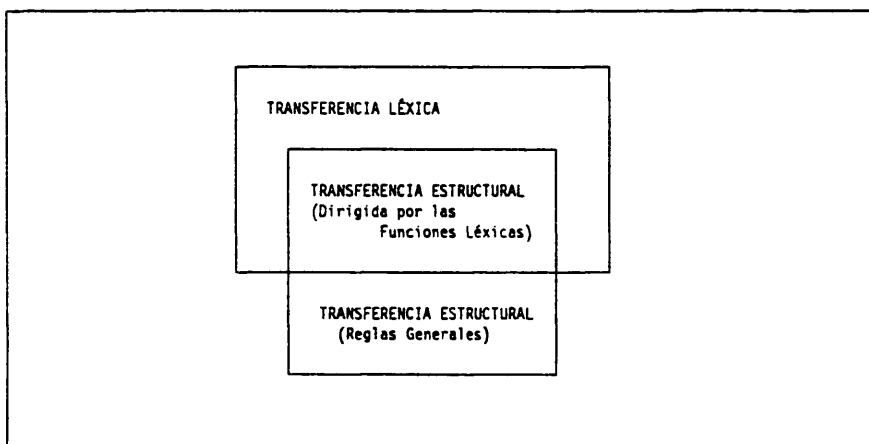


Figura 14. Relación entre las transferencias léxica y estructural. Transformaciones dirigidas por el léxico

La relación expuesta entre la transferencia léxica y estructural se representa en la Figura 14. En la Figura 6 en la página 37, donde se esquematizaba parte del proceso de traducción en MENTOR, aparecen la transferencia léxica y estructural organizadas secuencialmente con el fin de simplificar el gráfico, aunque la disposición real es la que se acaba de describir.

La decisión de permitir que el léxico provoque cambios estructurales nos ha ofrecido también la posibilidad de solucionar con elegancia toda una serie de problemas que se detectan en algunos análisis, debido a que PEG, diseñada en un principio para realizar únicamente análisis de la estructura superficial de las oraciones, no posee aún toda la información necesaria para poder distinguir estructuras sintácticas muy similares, y, como se verá en los siguientes capítulos, nos va a permitir evolucionar a medida que evolucione la gramática sin que la información léxica almacenada se vea afectada por los posibles cambios. Tenemos por ejemplo las siguientes oraciones:

- (2-5) I want him to go to school.
(2-6) I want a bicycle to go to school.

cuyas traducciones son:

- (2-5t) (Yo) quiero que (él) vaya a la escuela.
(2-6t) (Yo) quiero una bicicleta para ir a la escuela.

aunque en (2-6) se podría objetar una cierta ambigüedad, ésta no es muy probable (traducirlo como "(Yo) quiero que una bicicleta vaya a la escuela").

En el caso (2-5) se requiere una transformación estructural de la oración subordinada, "*him to go*", para poder realizar la traducción; pasando de ser una oración de infinitivo a ser una oración completiva con un verbo en forma personal, y cuyo sujeto equivale al pronombre que premodificaba al verbo de la oración inglesa. En el caso (2-6), que es formalmente idéntico, se presenta una situación completamente distinta debido a las características del NP que premodifica al verbo, "*a bicycle*".

DECL1	NP1	PRON1*	"i"
	VERB1*	"want"	
	NP2	PRON2*	"him"
	?	INFCL1	INFTO1 "to"
		VERB2*	"go"
	PUNC1	"."	

DECL1	NP1	PRON1*	"i"
	VERB1*	"want"	
	NP2	DETP1	ADJ1* "a"
		NOUN1*	"bicycle"
	?	INFCL1	INFTO1 "to"
		VERB2*	"go"
	PUNC1	"."	

Figura 15. Ejemplo de análisis ambiguos. En ambos casos se obtiene el mismo análisis

En principio no tendría por qué haber ninguna dificultad para distinguir los casos (2-5) y (2-6), ya que el análisis sintáctico es teóricamente distinto. En (2-6), el complemento directo del verbo "*want*" es todo lo que sigue, "*him to go*", mientras que en (2-6) es únicamente "*a bicycle*", actuando la oración de infinitivo únicamente como un indicativo de finalidad; pero para PEG ambos análisis son, en principio, ambiguos, como se representa en los árboles correspondientes (Figura 15).

La solución que hemos adoptado consiste en discriminar ambos casos considerando las características del objeto en la fase de transferencia léxica. Por supuesto que, cuando PEG posea el conocimiento semántico suficiente para desambiguarlos, esto no será necesario. Mientras tanto, sólo supone la inclusión de algunas indicaciones en la entrada, que, como se verá más adelante, pueden ser eliminadas en cualquier momento, sin ningún tipo de problemas marginales y sin afectar al funcionamiento global de la transferencia léxica. Además, tal como está diseñando el diccionario, aunque se dispusiera en algún momento de los análisis correctos, la operación de transferencia seguiría funcionando aún sin modificar la entrada antigua.

Después del acceso al diccionario, cuando se ha determinado ya la traducción de cada palabra, ésta se inserta como un valor más en los nodos del árbol decorado, en un campo específico que denominamos TBASE (Translated BASE). El contenido de dicho campo es la forma base de la palabra traducida, sin flexionar morfológicamente.

Como resultado práctico de nuestra experiencia en la implementación de la fase de transferencia, es necesario subrayar que, según ha ido enriqueciéndose la transferencia léxica y el diccionario bilingüe se ha ido haciendo más potente, muchas de las reglas que en un principio se habían considerado estructurales, o que se habían escrito para manejar problemas muy particulares porque no existía aún el mecanismo adecuado para hacerlo, fueron eliminándose, de forma que actualmente su número ha disminuido considerablemente, a la vez que el proceso de transferencia léxica se hace más y más importante.

2.2.3 La Generación

La fase de generación sólo se ocupa de temas relativos al español como lengua meta, y termina con la producción de la oración final traducida. Uno de los principales problemas de esta fase es la flexión morfológica de las palabras, que aparecen en su forma base después de la transferencia léxica, pero también hay otros problemas igualmente importantes como son el tratamiento de los pronombres clíticos, el orden de los complementos, las contracciones, los apócope, etc.

Por el momento no hemos incluido una gramática de generación propiamente dicha para la gestión de esta fase, aunque es uno de nuestros próximos objetivos. Lo que sí se han implementado son un conjunto de reglas que llevan a cabo la generación de la oración final, y que constituyen el núcleo inicial de dicha gramática.

TBASE, del que ya hablamos en el apartado anterior, es el atributo que se añade a los nodos terminales y a algunos núcleos, y cuyo valor es la traducción asignada. Por lo tanto, estos nodos son los que han de ser flexionados, incluyéndose la forma flexionada en el campo TSTR de dichos nodos (Translated STRing).

Para esto utilizamos el diccionario español LEXIS, desarrollado previamente en el Centro de Investigación UAM-IBM, del que se habla en el próximo apartado. Este diccionario se mantiene como un módulo externo, al igual que en el caso de la transferencia o el análisis; y es una función LISP, MORPHOLOGY, a través de un procedimiento específico, MORPHOLOGY-DRIVER, la encargada de establecer la comunicación con él, por medio de una interfaz ad hoc, escrita en REXX, que se adapta a nuestras necesidades. Las posibilidades que esta interfaz nos ofrece son:

1. **Análisis morfológico:** dada una forma flexionada, poder recuperar sus características morfológicas

(2-7) programas, NOMBRE → programa, NOMBRE PLURAL MASCULINO

2. **Generación morfológica:** dada una forma base y un conjunto de características morfológicas, poder recuperar la forma flexionada adecuada

(2-8) volver, VERBO PRESENTE INDICATIVO PERSONA3 SINGULAR → vuelve

Durante la generación nos interesa utilizar la segunda posibilidad, recuperando las formas flexionadas adecuadas dada la base de cada palabra en español. Las características necesarias para determinar esta forma se transmiten mediante un conjunto de parámetros que dependen de la parte de la oración de la palabra que se desea flexionar.

Como ya hemos dicho, en la fase de generación se tratan también otros problemas relacionados con la forma final de la oración traducida, que describimos brevemente a continuación.

2.2.3.1 Orden de los Complementos y Generación de Clíticos

El objeto indirecto, que en inglés puede preceder al directo, en español suele siempre aparecer después de éste, característica que hay que considerar en generación. Por otra parte, es posible también que haya que generar pronombres clíticos para adecuar la estructura superficial al modo español.

(2-9) I gave the writers the book.
(2-9t) (Yo) di el libro a los escritores.
(2-10) I gave the book to the writers.
(2-10t) (Yo) di el libro a los escritores.
(2-11) I gave them the book.
(2-11t) (Yo) les di el libro.
(2-12) I gave them it.
(2-12t) (Yo) se lo di.
(2-13) I gave it to the writers.
(2-13t) (Yo) se lo di a los escritores.

Como puede verse, en algún caso es incluso necesario duplicar el objeto indirecto, introduciendo en la oración un pronombre proclítico (2-13).

Estas operaciones suelen resultar bastante complejas porque muchas veces es necesario introducir nuevos nodos en la estructura del árbol y, casi siempre, reordenar los ya existentes (ver ejemplo detallado en la Figura 16 en la página 52).

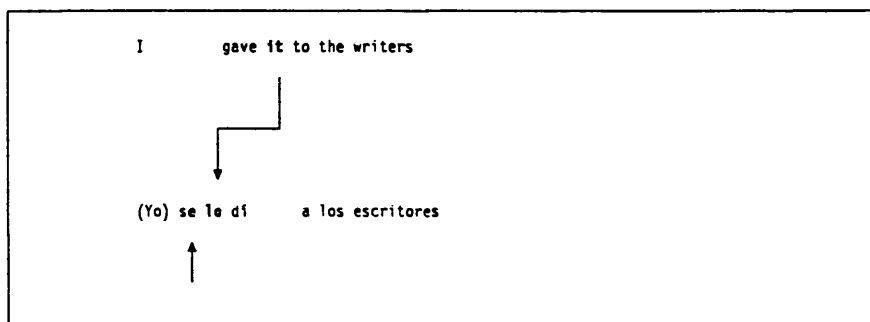


Figura 16. Duplicación y reordenación de pronombres. Fase de generación

También es necesario tener en cuenta otros aspectos laterales, como es la conservación de las concordancias en el género y número de los complementos pronominales que se incluyan.

2.2.3.2 Omisión de pronombres

Se han tratado también algunos casos en los que en español se omiten los pronombres. No se trata de eliminar los nodos correspondientes, sino de marcarlos de manera que aparezcan en la estructura, pero no se consideren al formarse la frase final. Por ejemplo, el pronombre "ello" se omite cuando el nodo correspondiente no es el tópico de la oración, entendido "tópico" en su sentido lingüístico, es decir, como el asunto o materia del que trata.

(2-14) It is easy to write a book.
(2-14t) Es fácil escribir un libro.

En (2-14) el tema es "to write a book" apareciendo el pronombre sujeto "it" como un sujeto podríamos decir auxiliar, vacío de contenido. En este tipo de construcciones en español se omite dicho sujeto, enunciándose la oración como en (2-14t).

2.2.3.3 Tratamiento de pasivas

En la fase de generación una de las primeras decisiones que hay que tomar es la relativa a las oraciones pasivas, ya que en español existe más de una posibilidad (pasiva refleja o normal). Todo esto unido al uso de los verbos "ser" y "estar", que constituye una gran diferencia del español respecto a otras lenguas. Tenemos, por tanto, dos aspectos a considerar en cuanto a la generación de oraciones pasivas:

- determinar si el verbo auxiliar equivale al español "ser" o al "estar"
- determinar si la pasiva ha de generarse como refleja o no

En los dos puntos anteriores se han tomado sendos criterios heurísticos que se van completando y perfeccionando, y que son los siguientes:

- en principio se genera "ser", salvo que se indique lo contrario
- todas las oraciones pasivas se generan como pasivas reflejas, salvo que se indique lo contrario.

Siendo el lugar en que se puede indicar lo contrario el diccionario bilingüe, como veremos en el próximo capítulo. Estos criterios, aún siendo aproximativos, nos han permitido obtener hasta ahora resultados altamente satisfactorios.

2.2.3.4 Presentación final de la oración

Una vez que ya han tenido lugar todas las transformaciones sobre el árbol, se recorren los nodos terminales y se construye la oración final concatenando las cadenas que se han ido generando (TSTR de los nodos). Llegados a este estadio, también es necesario muchas veces realizar algún tipo de operación "cosmética" para generar correctamente la oración final, sin afectar para nada a la estructura del árbol. Algunas de las operaciones posibles son:

- **Contracciones:** la concatenación de las preposiciones "a" o "de" con el artículo definido "el" da lugar a los artículos contractos "al" y "del" respectivamente, y son estas formas las que han de utilizarse en la oración final de salida.
- **Apócope:** algunos adjetivos, cuando preceden al nombre pierden sus última o dos últimas letras, teniendo que aparecer apocopados en la oración final. Por ejemplo:
"bueno" → "buen", "malo" → "mal", "grande" → "gran".
- **Pronombres clíticos:** que han de unirse al verbo siempre que éste les preceda:
"da" "le" "los" → "dáselos".

2.2.4 Los Diccionarios

Como ya se ha dicho, en cada una de las fases del proceso de traducción (análisis, transferencia y síntesis) se utilizan diccionarios específicos, formalmente distintos. En este apartado vamos a hablar de los diccionarios que se están utilizando actualmente en el análisis y la síntesis, introduciendo el modelo de diccionario utilizado en la fase de transferencia. En el siguiente capítulo se expondrán con detalle las características del formalismo en el que están escritas las entradas de este último diccionario.

2.2.4.1 Diccionario de análisis

En el análisis PEG hace uso de la información contenida en OD, Optimized Dictionary, información relativa a las palabras inglesas y que proporciona a PEG un vocabulario en principio ilimitado [Byr83]. Dicho diccionario se ha construido extrayendo información de los diccionarios *Webster's 7* [Web87] y *Longman* [Lon78].

En la Figura 17 se puede ver el aspecto que presentan las entradas de este diccionario cuando son consultadas por el usuario.

Cada palabra que se presenta al sistema durante el análisis es enviada al diccionario, que proporciona la siguiente información:

1. Las partes de la oración que esa palabra puede tomar, considerando todos sus sentidos.
Ej: want NOUN - want VERB, bicycle NOUN - bicycle VERB.
2. Información morfológica (número, identificación del tiempo verbal, etc.).
Ej: i PERS1 SING.
3. Características sintáctico-semánticas propias de la palabra, tales como el tipo de complemento que utiliza un verbo o la caracterización de un nombre como indicador de ente animado.
Ej: i ANIM HUM.

```
I
i (PRON ANIM HUM NOM PERS1 SING )

want
want(NOUN SING )
want(VERB INF NPING NPTOV NPTPT PLUR PRES TOVCOMP TRAN (OBJTPREP for) )

bicycle
bicycle(NOUN SING )
bicycle(VERB INF PLUR PRES )
```

Figura 17. Algunas entradas en el diccionario de análisis OD

Relacionados con este último aspecto se han incluido dos rasgos asociados a los verbos y que suponen un gran avance en la caracterización de éstos. Son OBJTPREP y PTC. OBJTPREP es un atributo que se asocia a los verbos que aparecen con complementos preposicionales concretos (preposición + NP), y cuyo significado depende la ocurrencia de dichos complementos. Es el

caso de *want* (OBJTPREP *for*) que se señala en el ejemplo. Aunque aquí sólo aparece una preposición, el caso general es que en OBJTPREP se proporciona una lista con todas las posibles.

De forma similar se incluye PTC en los verbos adecuados, listando esta vez las partículas que pueden aparecer asociadas al verbo cuando éste actúa como frasal (para aclaraciones sobre el concepto de verbos frasales y preposicionales, nos remitimos a las páginas 64 y 71).

La información contenida en OD es utilizada por el analizador para determinar cuál es la estructura correcta en cada caso y crear así los grafos de análisis y, posteriormente, los árboles.

OD se acompaña de una serie de programas auxiliares de acceso que permiten al usuario crear, borrar y modificar los rasgos de las entradas ya existentes. Para nuestra aplicación, traducción automática multilingüe, hemos creído que esta labor, aunque se puede realizar provisionalmente en los distintos grupos, debe centralizarse de manera que tanto el analizador como el diccionario del lenguaje fuente utilizado sigan siendo comunes para todos los lenguajes involucrados en el sistema de traducción.

Al igual que PEG, OD ha sido desarrollado en el laboratorio de IBM T.J. Watson Research Center de Yorktown Heights, Nueva York, y ambos están en un proceso continuo de enriquecimiento, que contempla la inclusión de más rasgos semánticos sobre los que basar los procesos de desambiguación en los casos conflictivos.

2.2.4.2 Diccionario de generación

En cuanto al diccionario que estamos utilizando para la síntesis, ya se ha mencionado que está basado en los datos recopilados por nuestro grupo para otros proyectos. Contiene más de 35.000 formas base que una vez flexionadas dan lugar a más de 400.000 palabras, y cubre ampliamente lo que constituye el lenguaje escrito habitual. Sin embargo, no contiene vocabulario técnico o especializado. La información que contiene es principalmente de tipo morfológico: el léxico asociado a cada base forma un conjunto de clases de inflexión o paradigmas, que indican cómo se flexionan las palabras del grupo, y caracteriza cada forma derivada con las características usuales de género, número, tiempo, persona, etc. Este diccionario está también en un proceso de enriquecimiento que incluirá la adición de códigos relativos a la subcategorización de las palabras de las distintas partes de la oración [Mag89].

2.2.4.3 Diccionario de transferencia

Para la fase de transferencia también es necesario utilizar un diccionario, o diccionarios, específicos. En nuestro prototipo decidimos que la información contenida en éstos habría de ser puramente bilingüe y sólo relacionada con el par de lenguas involucrado en el proceso.

La primera consideración que hay que tener en cuenta en el diseño de estos diccionarios es el tipo de palabras que van a contener. En este sentido se puede considerar que existirán muchas pala-

bras que se traducen directamente, al menos en el contexto técnico-informático que hemos considerado. Por otra parte, habrá algunas que sean específicas de nuestro dominio, y cuya traducción varíe completamente al cambiar el tema de los textos a traducir. Para terminar, existirá otro conjunto de palabras cuya traducción será muy compleja, y fuertemente dependiente del contexto sintáctico-semántico de cada ocurrencia. Por ello, independientemente de la estructura interna de los diccionarios bilingües, definimos una estructura de diccionarios multinivel, cuyo mecanismo de consulta permite la incorporación al sistema de cualquier diccionario adicional, sin necesidad de reestructurar el esquema general. La estructura diseñada consta de los siguientes diccionarios:

- **Diccionarios específicos:** dependientes de la aplicación concreta para la que se vaya a realizar la traducción, de un área técnica determinada, o bien que reflejen el sublenguaje específico de una entidad o empresa. Podrían existir, teóricamente, en número ilimitado, y permitirían adaptar el sistema a cada trabajo en concreto o a cada entidad, manteniendo el mismo núcleo básico formado por los programas de traducción y los dos diccionarios que siguen a continuación.
- **Un diccionario inteligente o especializado:** esto es, un diccionario en el que se recoge la información necesaria, tanto de la lengua origen como de la lengua meta, para poder determinar correctamente el sentido de las palabras con más de una traducción posible, así como las acciones peculiares que la elección de una traducción obligue a efectuar sobre la estructura de la oración de cara a la fase de generación.
- **Un diccionario de referencia:** en el que se almacenan las entradas simples (uno-a-uno).

En general, cuanto más específico sea un diccionario, antes deberá ser consultado.

El primero y el último diccionario mencionados poseen una estructura interna bastante sencilla, donde se indica simplemente cuál es la traducción adjudicada a cada entrada. Son entradas como la de *"table"*, para cuya traducción se ha escogido *"tabla"*, ya que es un vocablo válido en los contextos en que aparece en el dominio considerado.

Las entradas del diccionario especializado son, sin embargo, mucho más complejas, y su definición ha requerido el desarrollo de una metodología basada en el formalismo que se expone en la presente memoria. Presentamos también, a modo de ilustración, una de las entradas consideradas complejas, cuyas características y las del formalismo utilizado se detallan en los siguientes capítulos. Es la entrada del verbo *"play"*.

table NOUN

"tabla"

play VERB

```
($
+ ((F-OBJ)
($
+ ((M-IS '(part "cometido")) "desempeñar")
+ ((M-FEAT 'SPRT) (S-FCHANGE '((OBJ PREPC "a"))) "jugar")
+ ((M-FEAT 'HUM) (S-FCHANGE '((OBJ PREPC "con"))) "jugar")
=> ("tocar")))
+ ((F-CC 'PP)
($
+ ((M-FEAT 'INST) "jugar")
=> ("tocar")))
=> ("tocar"))
```

Actualmente, en nuestro sistema se utiliza: 1) un diccionario de referencia, de tamaño medio (unas 11.000 entradas) y 2) un diccionario especializado, específicamente diseñado para el prototipo, que hemos mencionado y del que hablaremos a continuación. Este último contiene actualmente unas 1.200 entradas.

2.3 Transferencia Léxica para una palabra: el verbo MAKE

En este apartado vamos a estudiar todo el proceso de la transferencia léxica para una palabra en concreto. Comenzaremos examinando la información que hay disponible a la hora de determinar sus distintas traducciones en función de los sentidos que ésta posee en la lengua origen. De la determinación de una traducción en concreto es asimismo posible que se deriven una serie de acciones sobre la estructura de la frase, y que tienen que ver exclusivamente con las características sintácticas y semánticas de la lengua meta. Estas implicaciones también han de determinarse simultáneamente con las traducciones.

Por último, se presentarán los resultados obtenidos sobre la TL de una manera ordenada, aproximando la información léxica a estructuras más computacionales.

La transferencia sobre la que nos centramos tiene al idioma inglés como lengua origen, y al español como lengua meta, y para este ejemplo se ha elegido un verbo, el verbo inglés "make", ya que en los verbos se centran algunos de los problemas más importantes de la TL, al ser una de las partes de la oración más complejas. Además, su traducción muchas veces afecta a la propia TL de otros constituyentes más o menos directamente relacionados con éstos.

Para llevar a cabo la extracción de información que efectuamos en las páginas siguientes, hemos determinado una serie de reglas de operación que nos permiten establecer ciertos criterios a la hora de determinar las traducciones:

1. En caso de varias traducciones posibles para un mismo sentido, que resulten ser sinónimos, se elegirá la traducción más general, que pueda ser aplicada en un mayor número de ocasiones. De forma que no se considerarán, por ejemplo, las expresiones familiares.
2. Se desecharán las acepciones que probablemente no vayan a darse en nuestro dominio de trabajo, a no ser que introduzcan algún problema especialmente interesante. Nuestro dominio, como se ha dicho en otros apartados del capítulo, es el de los textos técnicos en el entorno de la informática.
3. Para una misma acepción, si todas las traducciones que se proporcionan son análogas, entonces se elegirá sólo una.
4. Por último, se garantizará que se va a obtener una traducción en cualquier caso; existiendo las llamadas 'traducciones por defecto'. Las traducciones por defecto se asignarán si no se dan ninguna de las condiciones previstas en cada caso.

2.3.1 Información disponible

Para hacer un estudio de este tipo, la principal fuente de información de que se dispone son los diccionarios bilingües. La elaboración de un buen diccionario (mono, bi, o multilingüe) es siempre una tarea compleja y, si lo que se busca no es un instrumento de bolsillo, el número de buenos diccionarios disponibles es limitado. En el caso concreto del par de lenguas inglés-español, entre los que más se usan actualmente se encuentran los diccionarios Collins [Col84] y Larousse [Lar84].

Respecto al método de trabajo, en principio hay dos opciones:

1. extraer la información de la entrada elegida de varios diccionarios, con la intención de cotejar los distintos contenidos y llegar a una definición más rica
2. elegir un único diccionario y formalizar la información que éste nos proporciona, creando una estructura básica que se pueda completar más tarde

La primera opción puede parecer la más correcta, pero esto no es tan evidente dadas las características del tipo de información considerada y de las que ya hemos hablado en el Capítulo 1; dicha información está muy poco formalizada y tampoco tiene una forma descriptiva estándar, por lo tanto, dependiendo del diccionario, los sentidos que se proporcionen pueden no coincidir, o pueden incluso no ser exactamente los mismos. En la Figura 18 en la página 61 se muestra un ejemplo presentado por O. Rizk [Riz89]. En él se relacionan los sentidos adjudicados a la misma palabra inglesa, "mug", en tres diccionarios diferentes: *Webster Seventh Collegiate Dictionary (W7)*, *Longman's Dictionary of Contemporary English (LDOCE)* y *Collins English-French*.

En el diccionario inglés-francés, el nombre "mug" tiene tres sentidos. El sentido a) es el de un taza grande o jarra; el sentido b) se refiere al rostro, pero con matiz familiar (jeta o morro); y el sentido c) es el de bobo o tonto, refiriéndose a una persona. Estos tres sentidos se distinguen en los dos diccionarios monolingües W7 y LDOCE, aunque LDOCE añade otro sentido a a), como el contenido mismo de la jarra. W7, por su parte, añade un sentido 2c como un tipo de gráfico o pintura, que puede considerarse como una extensión de b). Y así sucesivamente.

Este ejemplo nos da idea de lo complejo que resulta la determinación de las equivalencias entre los diferentes sentidos que proporcionan distintos diccionarios, ya que no sólo se trata de formalizar la TL para una palabra a partir de la información disponible, sino que dicha información ha de ser previamente re-estructurada para poder asegurar su coherencia.

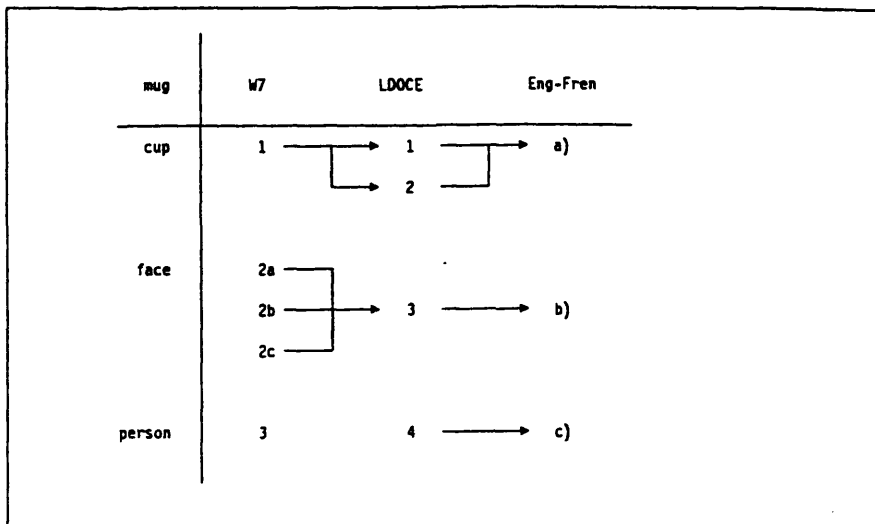


Figura 18. Definiciones en distintos diccionarios. Los sentidos no siempre coinciden.

Por otra parte, desde el punto de vista computacional, para las aplicaciones que se quieren implementar (traducción de textos técnicos) no es necesario que se posea una descripción exhaustiva de cada entrada.

Como conclusión, hemos decidido comenzar el trabajo como se señalaba en segundo lugar: eligiendo un único diccionario y formalizando la información que éste nos proporciona, con el fin de crear una estructura básica que se pueda completar más tarde.

En nuestro grupo disponemos en línea del Collins inglés-español [Col84], y es éste el que tomaremos como referencia.

2.3.2 Extracción de la Información

En las siguientes páginas se puede ver la entrada completa que aparece para el verbo inglés "make" en el Collins inglés - español. A partir de esta información, de tipo textual, vamos a extraer las características formales de las acepciones que se consideren, que no serán todas, puesto que nos ceñiremos sólo a las que puedan ocurrir en un entorno técnico o, en algún caso, a las que tengan un interés especial por razones de tipo lingüístico o formal.

make

[meik] (irr: pret and ptp made)

1 vt

- (a) (general sense) hacer; fabricar; construir; elaborar; formar; crear; componer; bed, effort, fire, noise, peace, remark, tea, war, will hacer; dress confeccionar; meal preparar; speech pronunciar; error cometer; payment efectuar; cards barajar; face poner; sense tener; to someone a judge constituir a uno juez, nombrar a uno juez; to someone king elevar a uno al trono; they've made John secretary han puesto a Juan por secretario; to a friend of someone trabar amistad con uno; he's as cunning as they 'em es de lo más astuto, es sumamente astuto; I'm not made for running yo no estoy hecho para correr; to be made of estar hecho de, estar compuesto de, consistir en, constar de; it's made of gold es de oro, está hecho de oro; to show what one is made of demostrar las cualidades que tiene uno; what do you _ of this? ¿qué te parece esto?; what did you _ of the film? ¿qué impresión te produjo la película?; what do you _ of him? ¿qué piensas de él?; ¿qué impresión te has formado de él?; I can _ nothing of it no lo entiendo, no saco nada en claro; I don't know what to _ of it no me lo explico.
- (b) (complete, constitute) circuit cerrar; trick ganar, hacer; 2 and 2 _ 4 2 y 2 son 4; that _s 20 con éste hacen 20; it still doesn't _ a set todavía no completa una serie entera; it doesn't _ a full course no equivale a una asignatura completa, no es igual a una serie completa; to _ a contract (Cards) cumplir un contrato; South leads and _s 5 tricks Sur sale y efectúa 5 bazas; it made a nice surprise fue una sorpresa agradable; partridges _ good eating las perdices son buenas de comer; it _s pleasant reading da gusto leerlo; he made a good husband resultó ser un buen marido; he'll _ a good footballer será buen futbolista, tiene madera de futbolista; I made one of the party yo era (uno) del grupo.
- (c) (earn etc) ganar; he _s £30 a week gana 30 libras a la semana; how much do you _? ¿cuánto ganas?; ¿qué sueldo cobras?; to _ a fortune enriquecerse, hacer su pacotilla (fam); what will you _ by it? ¿cuánto vas a ganar en esto?; how much do you stand to _? ¿cuánto esperas ganar?

- (d) (assure future of) hacer la fortuna de; asegurar el triunfo de; this film made her esta película fue el principio de su éxito; he was made for life se aseguró un porvenir brillante; to _ or break someone hacer la fortuna o ser la ruina de uno; to _ or war something decidir de una vez la suerte de algo.
- (e) (with pred adj) hacer; to _ someone happy hacer a uno feliz; to _ someone angry irritar a uno, provocar a uno, sacar a uno de quicio; to _ someone ashamed dar vergüenza a uno; to _ someone sleepy dar a uno ganas de dormirse; to _ someone rich enriquecer a uno; to _ someone ill sentar a uno mal; to _ something ready preparar algo; to _ iron hot calentar un trozo de hierro; to _ one's voice heard hacer que se escuche la voz de uno.
- (f) (say, agree) let's _ it 9 o'clock citémonos para las 9, pongamos las 9.
- (g) (believe) creer; the situation is not so bad as you _ it la situación es menos grave de lo que Vd cree.
- (h) (calculate) what do you _ the time? ¿qué hora tienes?; I _ it 7.30 yo tengo las 7 y media; how many do you _ it? ¿cuántos dices tú?; ¿cuántos tienes en total?
- (i) (force) to _ someone do something forzar (or obligar, compeler) a uno a hacer algo; (persuade) inclinar (or inducir) a uno a hacer algo; you can't _ me Vd no puede forzarme a hacerlo; what _s you do it? ¿por qué te ves obligado a hacerlo?; what made you say that? ¿por qué dijiste eso?; to _ someone laugh mover a uno a risa.
- (j) (reach, attain) we made 15 knots alcanzamos una velocidad de 15 nudos; we shall never _ the shore no llegamos nunca a la playa, será imposible alcanzar la playa; to _ it (arrive) llegar; (achieve something) conseguir lo que se deseaba; (succeed) tener éxito, triunfar; eventually we made it por fin llegamos; we just made it in time llegamos apenas con el tiempo justo; can you _ it by 10? ¿puedes llegar para las 10?; to _ it with someone (sl) conseguir acostarse con una.
- (k) (with adv or prep) to _ into convertir en, transformar en; to _ out cheque, document, receipt extender; list hacer, redactar; form llenar; the cheque should be made out to Pérez el cheque será nominativo a favor de Pérez; to _ out (see) distinguir,

vislumbrar, divisar; writing descifrar;
(understand) entender; I can't _ it out at
 all no me lo explico, no lo entiendo; he
 made out that... dio a entender que... , dio
 la impresión de que... , nos hizo creer
 que... ; you _ him out to be better than he
 is Vd hace creer que es mejor de lo que es
 en realidad; he's not as rich as people _
 out es menos rico de lo que dice la gente;
 how do you _ that out? cómo deduces eso?,
 ¿de qué se infiere eso?; to _ over ceder,
 traspasar; to _ up (make) hacer; fabricar;
dress confeccionar; medicine preparar;
collection reunir; parcel empaquetar;
 envolver; list hacer, redactar; print, page
 componer; to _ up (compose) componer,
 formar, constituir; the parts which _ it up
 las partes que lo componen; the party was
 made up of 8 bishops el grupo lo integraban
 8 obispos; to _ up (invent) inventar; you're
 making it up! ¡puro cuento!; to _ up
(complete) completar; quantity, total
 completar, hacer; to _ up loss subsanar;
lost time etc recuperar; deficit cubrir; to
 _ it up to someone compensar a uno,
 indemnizar a uno; to _ up a fire echar
 carbón a una lumbre; to _ up a quarrel hacer
 las paces; to _ it up with someone hacer las
 paces con uno; to _ up one's face pintarse,
 maquillarse; to _ up an actor maquillar a un
 actor.

- 2 vi (a) (tide) crecer, subir.
 (b) he made as if to + infin hizo como si
 quisiese + infin, fingió que iba a + infin,
 hizo además de + infin.
 (c) to _ away, to _ off largarse; huir,
 escaparse; to _ away with, to _ off with
 llevarse, alzarse con; escaparse con; to _
 away with someone matar a uno; to _ away

- with oneself quitarse la vida, suicidarse.
 (d) to _ after someone seguir a uno,
 perseguir a uno; to _ for a place, to _
 towards a place dirigirse a un lugar,
 encaminarse a un lugar; where are you making
 for? ¿adónde se dirige Vd?; to _ for someone
 atacar a uno, abalanzarse sobre uno.
 (e) to _ for result etc contribuir a,
 conducir a; it _s for optimism ayuda a crear
 el optimismo, fomenta el optimismo; it _s
 for difficulties tiende a crear
 dificultades.
 (f) to _ out arreglárselas, salir bien;
 we're making out vamos tirando; we made out
 eventually por fin nos las arreglamos; how
 are you making out? cómo te va eso?
 (g) to _ up pintarse, maquillarse.
 (h) to _ up for someone's losses compensar a
 uno por sus pérdidas, indemnizar a uno de
 sus pérdidas; to _ up for a lack of supply
 una falta de; to _ up for lost time
 recuperar el tiempo perdido.
 (i) to _ up to someone (procurar)
 congratarse con uno, (procurar) ganarse la
 amistad de uno; halagar a uno, hacer
 zalamerías a uno.

3 vr:

- (a) to _ oneself an expert in llegar a ser
 experto en; to _ oneself dictator hacerse
 dictador, constituirse en dictador.
 (b) to _ oneself comfortable acomodarse a su
 gusto; to _ oneself ill with work enfermar
 por exceso de trabajo; to _ oneself
 ridiculous ponerse en ridículo; see hear
etc.
 (c) to _ oneself do something obligarse a
 hacer algo; I have to _ myself (do it) tengo
 que hacer un esfuerzo (por hacerlo).

En primer lugar aparece la transcripción fonética del verbo, así como cierta información morfológica sobre su comportamiento como irregular. Esta información no es relevante en nuestro caso.

La primera división que se observa en la entrada está indicada por los sangrados que, numerados a la izquierda del 1 al 3, identifican los distintos usos funcionales del verbo: transitivo, vi, intransitivo, vi, y reflexivo, vr, y dan lugar a tres apartados diferentes.

Cada uno de estos apartados de los que hablamos está a su vez dividido, esta vez alfabéticamente. Estas divisiones determinan distintas acepciones para el uso del verbo en cada caso. Por

último aparecen ya las traducciones en sí, generalmente acompañadas de ejemplos que facilitan su identificación. La palabra objeto de la entrada ("make") se representa mediante un signo característico, , por razones de claridad.

Podemos ver que el tipo de letra también resulta ser un modo de codificación de la información. Así, entre paréntesis y subrayadas, aparecen a veces algunas palabras que identifican el contexto en el que se obtiene la traducción indicada. Por otra parte, se presentan en negrita los ejemplos de las traducciones posibles para cada acepción, que no siempre corresponden a una forma en infinitivo, sino que también pueden ser oraciones completas, sobre todo cuando no es simple indicar en qué contexto se obtiene esa traducción, o cuando se trata de frases hechas.

Como se puede observar, y como ya hemos comentado en apartados anteriores, ésta no es información codificada, aunque para un lexicógrafo estas imprecisiones o carencias de información son solventables.

Vamos a seguir ahora analizando manualmente la entrada que nos ocupa, el verbo "make". Primero tenemos las tres grandes divisiones que nos presenta el Collins; éstas obedecen a características funcionales del verbo inglés, y son las siguientes:

1. vt para *verbo transitivo*, donde aparece un objeto directo que recibe la acción del verbo,
2. vi para *verbo intransitivo*, donde ningún objeto recibe la acción,
3. vr para *verbo reflexivo*, donde el objeto que recibe la acción del verbo es el propio agente.

El tercer caso se puede considerar como una ocurrencia particular del primero, en el que el objeto coincide con el sujeto. Como para nuestros fines no tiene especial interés mantener esta diferenciación, consideraremos en el mismo grupo los puntos 1 y 3.

Pasamos a analizar cada uno de estos dos grupos por separado, teniendo en cuenta que no todos los sentidos que detectemos nos serán de utilidad (p.e. los usos literarios o familiares), de manera que, como ya hemos dicho, sólo extraeremos aquellos que en principio parezca que puedan utilizarse en un entorno científico-técnico o resulten significativos o especialmente interesantes. Como este método de trabajo es incremental, una vez formalizada la información, siempre cabe la posibilidad de incluir nuevas acepciones.

2.3.2.1 Acepciones Transitivas

La principal característica de los usos transitivos es la existencia de un constituyente muy íntimamente ligado al verbo, como es el objeto directo. Como ya veremos, no sólo su presencia será importante, sino que también pueden serlo la presencia o no de preposiciones e incluso las características que pueden diferenciar un tipo de objeto directo de otro.

En inglés, los verbos transitivos pueden ser de tres tipos:

- **frasales (phrasal verbs):** se requiere una partícula antes o después del objeto directo,
- **preposicionales (prepositional verbs):** la función de objeto directo es asumida por un complemento preposicional,
- **simples:** el objeto directo aparece como un NP simple, sin partículas de ningún tipo.

De cara a nuestra formalización, los dos primeros casos pueden ser asimilables, de manera que sólo se considere si el objeto aparece como un NP simple o no.

Hemos extraído un par de acepciones del grupo (k) (with adv or prep) que ejemplifiquen los primeros casos:

- (vt-k2) *"to make out"*, aunque esta acepción es enormemente rica y compleja, hemos elegido una única traducción, que se tomará por defecto, y se cree que es la más probable en el entorno considerado: *"distinguir"*.
- (vt-k4) *"to make up"*, este caso también es muy rico en traducciones posibles, pero sólo vamos a elegir, como más probable, su traducción en el sentido de *"constituir"* o *"componer"*.

Esta primera aproximación a los casos transitivos queda entonces:

```
si OBJETOP con (OUT => "")
(vt-k2) => "distinguir"
si OBJETOP con (UP => "")
(vt-k4) => "componer"
si no
(vt-DEF) => "hacer"
```

Figura 19. Formalización de acepciones (1). MAKE

Donde se ha denotado 'OBJETOP' a los casos transitivos (luego tienen objeto directo) y preposicionales o frasales, de ahí la 'P' que añadimos a la palabra 'OBJETO'.

En los caracteres entre paréntesis que aparecen a la izquierda de cada acepción se señala el uso del verbo (vt, para verbo transitivo), la letra correspondiente a la acepción según la clasificación del Collins (a, b, etc.), así como un número que se asigna correlativamente a las diferentes traducciones posibles dentro de cada acepción. Este código se mantendrá durante todo el desarrollo para mayor claridad. Además se ha incluido una última posibilidad de traducción, (v-DEF), que significa traducción "por defecto", y es la que se ha de obtener como traducción del verbo *"make"* siempre que no se de ninguno de los casos particulares considerados. Evidentemente, ésta traducción ha de ser la más general, que en nuestro caso corresponde a la acepción del grupo (a) (general sense): *"hacer"*.

En estas expresiones hemos utilizado un pseudocódigo en lenguaje natural, donde el símbolo final ' \Rightarrow ' indica la traducción asignada al verbo objeto de la transferencia léxica, "make", e indicará traducción en todos los casos. Por ejemplo, las partículas reseñadas están íntimamente ligadas al verbo y no tienen traducción por sí mismas, siendo esto es lo que se ha querido indicar asociándoles una traducción vacía "".

La primera aproximación que hemos realizado es esencialmente correcta, pero pensemos ahora cómo se traducirían las siguientes frases según el esquema de la entrada propuesta.

(2-15) These are the elements which make up the system.

(2-16) The group was made up of four people.

Con la primera oración no habrá ningún problema:

(2-15t) Estos son los elementos que constituyen el sistema.

pero la segunda oración, que en inglés se puede considerar pasiva, en español pasa a ser una perífrasis atributiva con el verbo "estar", sentido que se recoge en la acepción (a):

(2-16t) El grupo estaba compuesto por cuatro personas.

de forma que no entra en las reglas estructurales generales para la fase de generación de las oraciones pasivas (ver el apartado "2.2.3 La Generación" en la página 50). Esta situación es conveniente indicarla explícitamente: por una parte no ha de considerarse como un caso corriente de pasiva y, por la otra, ha de generarse el verbo "estar" como auxiliar, no el verbo "ser" como es usual.

Otro detalle que podemos observar es que la preposición que introduce el complemento preposicional que aparece como postmodificador en (2-16) ("of"), no se traduce por su valor más general, "de", sino que se traduce como "por". Esto también ha de indicarse de alguna manera. En resumen, incluyendo los casos anteriores, tenemos el esquema de la Figura 20 en la página 67.

Como casos tipo de ocurrencias transitivas sólo nos quedan ahora las ocurrencias transitivas simples: ni preposicionales ni frasales. De ellas, un caso bastante sencillo de identificar es el correspondiente al grupo (e) (with pred adj), en el que, además del objeto directo, existe un adjetivo con valor predicativo que se refiere a éste. Por ejemplo: "to make someone happy". En estos casos, la traducción del verbo suele ser "hacer", pero no siempre; consideremos "to make a program ready", que se traduce como "preparar un programa". Si no existieran excepciones, se podría intentar que este caso entrara en el conjunto de traducciones por defecto, asignando "hacer" al verbo pero, como hemos visto, puede darse el caso de que la traducción de éste varíe. Esta circunstancia se puede expresar como aparece en la siguiente figura, en la que por (PSMOOS = PREDADJ) se

indica la existencia de un postmodificador del verbo que es un adjetivo predicativo, de forma que si éste existe y es "ready" se obtiene la traducción (vt-e7) "preparar", adjudicándose la traducción por defecto (vt-e1) "hacer" en caso contrario.

```

    si OBJETOP con (OUT => "")
(vt-k2) => "distinguir"
    si OBJETOP con (UP => "") y
    si es PASIVA
(vt-k4) => "componer" + NO ACTIVA + ESTÁTICA + si tiene (PP con (OF => "por"))
    si no
(vt-k4') => "constituir"
    si no
(vt-DEF) => "hacer"

```

Figura 20. Formalización de acepciones (2). MAKE

```

    si OBJETO y
    si tiene (PSMODS = PREDADJ) y
    si (PREDADJ es (READY => ""))
(vt-e7) => "preparar"
    si no
(vt-e1) => "hacer"
    si no
(vt-DEF) => "hacer"
    si OBJETOP con (OUT => "")
(vt-k2) => "distinguir"
    si OBJETOP con (UP => "") y
    si es PASIVA
(vt-k4) => "componer" + NO ACTIVA + ESTÁTICA + si tiene (PP con (OF => "por"))
    si no
(vt-k4') => "constituir"
    si no
(vt-DEF) => "hacer"

```

Figura 21. Formalización de acepciones (3). MAKE

El apartado (I) también es similar en cuanto a que la forma del complemento ayuda a determinar la traducción. Aquí lo que se identifica es una oración de infinitivo, que llamaremos INFCL (INFiniteive CLause), actuando como postmodificador:

- (2-17) The circumstances made him write a new book.
 (2-17t) Las circunstancias le obligaron a escribir un nuevo libro.

La traducción de "make" en este caso es "obligar a", traducción que introduce un nuevo problema: la expresión en la lengua meta de traducciones que no sean una única palabra, sino una fórmula. Las fórmulas están compuestas de dos o varias palabras, de las cuales no todas se flexionan morfológicamente (p.e. "salir bien", "dar una vuelta"). En nuestro proceso de formalización, un método posible para resolver estos casos es el siguiente: 1) permitir que una palabra, la entrada, se traduzca por una fórmula y 2) especificar, para dicha fórmula, cuáles de sus palabras se flexionarán morfológicamente y cuáles no. El segundo punto puede realizarse codificando la información correspondiente de alguna manera, por ejemplo, asignando un '0' a la o las palabras de la fórmula que no se flexionen, y la parte de la oración que les corresponde a las que sí se flexionan: 'VERB', 'ADJ', etc.

Por otra parte, en (2-17) el infinitivo de la oración subordinada ("write a new book") permanece como tal infinitivo en el proceso de transferencia, pero esta estabilidad no es general y, en principio, depende del verbo que gobierna la oración y de las características particulares de cada ocurrencia; basta comparar el ejemplo anterior (2-17) y su traducción con los ejemplos (2-5) y (2-6) y las suyas, que aparecían en este mismo capítulo:

(2-5) I want him to go to school.
(2-5t) (Yo) quiero que (él) vaya a la escuela.

(2-6) I want a bicycle to go to school.
(2-6t) (Yo) quiero una bicicleta para ir a la escuela.

Es por tanto razonable que en la entrada del verbo, en este caso "make", se proporcione también el tipo de transferencia que afecta a las oraciones de infinitivo que le complementen. Esta transferencia es de carácter estructural en cuanto a la lengua meta y la indicamos como tr-estructural en el esquema de la Figura 22 en la página 69.

Para terminar ya con este ejemplo hemos elegido dos traducciones que han sido extraídas de la acepción (a) y que introducen un caso interesante que aún no habíamos contemplado, como es el problema de la variación funcional de los constituyentes de la oración durante el proceso de traducción:

(2-18) You should make the previous function a new tool.

(2-19) They will make him president.

Ambas situaciones corresponden también a un tipo de transferencia que es discriminable en función de otros constituyentes, en este caso de la presencia de un segundo NP relacionado directamente con el objeto. Sin embargo, en estas dos oraciones el verbo tiene traducciones distintas:

(2-18t) Usted debería hacer de la función previa una nueva herramienta.

(2-19t) (Ellos) le nombrarán presidente.

```

    si OBJETO y
    si tiene (PSMODS = PREDADJ) y
    si (PREDADJ es (READY => ""))
(vt-e7)    => "preparar"
    si no
(vt-e1)    => "hacer"
    si tiene (PSMODS = INFCL)
(vt-i1)    => "obligar" "a" + flexionar (VERB 0) + tr-estructural (INFNP)
    si no
(vt-DEF)   => "hacer"
    si OBJETOP con (OUT => "")
(vt-k2)    => "distinguir"
    si OBJETOP con (UP => "") y
    si es PASIVA
(vt-k4)    => "componer" + NO ACTIVA + ESTÁTICA +
    si tiene (PP con (OF => "por"))
    si no
(vt-k4')   => "constituir"
    si no
(vt-DEF)   => "hacer"

```

Figura 22. Formalización de acepciones (4). MAKE

La causa de estas traducciones diferentes es la naturaleza del objeto directo: es humano en el segundo caso y en el primero no. Por otra parte, además de tener distintas traducciones, en el segundo caso las características de los constituyentes se mantienen durante la transferencia, mientras que en el primero la función de uno de ellos varía durante el proceso. Esta variación consiste en que el primer NP, "the previous function", que es el objeto en la lengua origen, pasa a ser un complemento preposicional en la lengua meta, introducido por la preposición "de". Esta transferencia funcional se indicará en la parte de acciones dependientes de la lengua meta, como puede verse en esta última figura, traducción (vt-a1), donde también se incluyen las traducciones por defecto asignadas.

```

    sí OBJETO y
    sí tiene (PSMODS = NP) y
    sí (NP es HUMANO)
(vt-a2)    => "nombrar"
    sí no
(vt-a1)    => "hacer" + tr-funcional (OBJETO -> PP "de")
    sí tiene (PSMODS = PREDADJ) y
    sí (PREDADJ es (READY => ""))
(vt-e7)    => "preparar"
    sí no
(vt-e1)    => "hacer"
    sí tiene (PSMODS = INFCL)
(vt-i1)    => "obligar" "a" + flexionar (VERB 0) + tr-estructural (INFNP)
    sí no
(vt-DEF)    => "hacer"
    sí OBJETOP con (OUT => "")
(vt-k2)    => "distinguir"
    sí OBJETOP con (UP => "") y
    sí es PASIVA
(vt-k4)    => "componer" + NO ACTIVA + ESTÁTICA +
    sí tiene (PP con (OF => "por"))
    sí no
(vt-k4')    => "constituir"
    sí no
(vt-DEF)    => "hacer"

```

Figura 23. Formalización de acepciones (5). MAKE

No consideramos el resto de las acepciones del uso transitivo ni el uso reflexivo del verbo "make", ya que no aportan ninguna situación de especial interés para nuestros propósitos.

2.3.2.2 Acepciones Intransitivas

En la entrada que estamos estudiando existen nueve acepciones diferentes para el uso intransitivo de "make", desde la (a) hasta la (i).

La acepción (a) no es la más general, puesto que se refiere a una traducción posible del verbo en el caso de que se esté hablando de (*tide*) ("marea"). Esta acepción no nos interesa, al menos por el momento, de modo que no la consideraremos. La (b) en principio no habría que considerarla si la traducción "por defecto" del verbo "make" se determina como "hacer" y se puede garantizar que es la que se va a obtener en este caso. De manera que, para no complicar innecesariamente la codificación, pospondremos la resolución de este apartado hasta concretar el resto de las traducciones.

De la acepción (c) a la (f) tenemos un conjunto de casos *frasales* y *preposicionales* con diferentes partículas y preposiciones.

Al igual que en el caso transitivo, los verbos intransitivos pueden ser simples, preposicionales y frasales. En los dos últimos casos la traducción del verbo va ligada a la presencia de la partícula (*frasal*) o complemento preposicional (*preposicional*), que han de estar presentes en la oración. Sin embargo, aún presentándose el mismo tipo de construcción para las acepciones transitivas e intransitivas en lo que a usos frasales y preposicionales se refiere, en estas últimas acepciones no resulta conveniente integrar en el mismo grupo los dos casos mencionados, como hicimos en el apartado anterior, ya que ahora no existe un complemento análogo al objeto directo que nos permita simplificar la estructura de las entradas. Por otra parte, veremos que en los casos intransitivos frasales la partícula asociada al verbo suele desaparecer al realizar la traducción, mientras que, en el otro caso, la traducción de las preposiciones muy a menudo viene determinada por la propia traducción del verbo. Es por ello que hemos decidido mantener las dos situaciones posibles como casos diferentes, utilizando la clave PTC para los usos frasales y la clave PP para los preposicionales. En ambos casos se especifica la partícula o preposición considerada, como en el apartado anterior.

De la acepción (c), que incluye usos frasales, omitimos las últimas traducciones ("*matar a uno*", "*suicidarse*"), que no es muy probable que aparezcan en el dominio en el que nos centramos. El resto podría expresarse:

```
si tiene (PTC = (AWAY => ""))
(vi-c1) => "escaparse"
si tiene (PTC = (OFF => ""))
(vi-c2) => "escaparse"
si no
(vi-DEF) => "hacer"
```

Figura 24. Formalización de acepciones (6). MAKE

En cuanto a la traducción asignada al verbo, en el texto vemos que se ofrecen varias posibles: "*largarse*", "*huir*" y "*escaparse*". De entre las tres, que se pueden tomar como sinónimos, hemos elegido la última, que resulta ser adecuada en un mayor número de contextos ("*largarse*", por ejemplo, se puede considerar como un uso coloquial). De todas maneras, la palabra "*escaparse*" en realidad está compuesta del infinitivo de un verbo más el pronombre clítico que se utiliza en su conjugación:

(2-20) "escapar" = "escapar" + "se"

Un ser humano entiende fácilmente esta expresión yuxtapuesta, pero para nuestros fines computacionales, resulta más conveniente señalar el caso de forma explícita, p.e. como se muestra a continuación.

```

    si tiene (PTC = (AWAY => ""))
(vi-c1)  => "escapar" + "se"
    si tiene (PTC = (OFF => ""))
(vi-c2)  => "escapar" + "se"
    si no
(vi-DEF) => "hacer"

```

Figura 25. Formalización de acepciones (7). MAKE

Pasemos ahora a las siguientes acepciones. En el grupo (d) tenemos tres posibilidades: *"to make after someone"*, *"to make for a place"* y *"to make towards a place"*, de las que se también se proporcionan ejemplos de uso. Todas ellas corresponden a construcciones preposicionales del verbo. Esta situación se puede formalizar indicando de alguna manera que se requiere un complemento preposicional introducido por una cierta preposición, como se indica en el esquema.

```

    si tiene (PTC = (AWAY => ""))
(vi-c1)  => "escapar" + "se"
    si tiene (PTC = (OFF => ""))
(vi-c2)  => "escapar" + "se"
    si tiene (PP = (AFTER => ""))
(vi-d1)  => "seguir"
    si tiene (PP = (FOR => ""))
(vi-d2)  => "dirigir" + "se"
    si tiene (PP = (TOWARDS => ""))
(vi-d3)  => "dirigir" + "se"
    si no
(vi-DEF) => "hacer"

```

Figura 26. Formalización de acepciones (8). MAKE

Pero tal como está expresado, la acepción para el uso preposicional *"to make for"* incluiría tanto el sentido de (d) como el que se indica en el apartado (e), *"to make for result"*. Por lo tanto es necesario establecer alguna diferencia entre ambas acepciones. La diferencia que existe tiene que ver con el tipo de complemento preposicional que aparece: en el primer caso es un lugar (*to make for a place*), mientras que en el segundo es una clase de *resultado*. En principio es más sencillo determinar si algo es o no un *lugar*, que determinar si puede o no ser un *resultado*, de forma que utilizaremos esta primera característica como elemento diferenciador.

En la siguiente figura podemos ver cómo queda ahora el esquema formal de la entrada, donde la marca PL se utiliza como distintivo de lugar (PLace), y se ha llamado NPP al núcleo objeto del complemento preposicional.

```

      si tiene (PTC = (AWAY => **))
(vi-c1) => "escapar" + "se"
      si tiene (PTC = (OFF => **))
(vi-c2) => "escapar" + "se"
      si tiene (PP = (AFTER => **))
(vi-d1) => "seguir"
      si tiene (PP = (FOR => **)) y
      si (NPP de tipo PL)
(vi-d2) => "dirigir" + "se"
      si no
(vi-e1) => "contribuir"
      si tiene (PP = (TOWARDS => **))
(vi-d3) => "dirigir" + "se"
      si no
(vi-DEF) => "hacer"
```

Figura 27. Formalización de acepciones (9). MAKE

Siguiendo con el mismo procedimiento, pasemos a analizar el caso (f). De entre las posibles traducciones que se sugieren para "to make out" elegimos "salir bien", que parece ser la más general, y resulta ser una fórmula análoga a la que se estudió en el apartado anterior para los casos transitivos: "obligar a". El resultado de formalizar la traducción en estos términos se puede ver en la figura de la página siguiente.

Quedarían por analizar las acepciones (g), (h) e (i), pero no vamos a incluirlas en este desarrollo porque no son representativas de otros casos diferentes a los ya analizados. La acepción (b) podemos ver que queda resuelta con la estructura que hemos propuesto.

En el caso de "make" se puede observar que el tratamiento de las acepciones intransitivas es más simple que el de las transitivas, esto es debido principalmente a que no se tiene que tener en cuenta el objeto directo para diferenciar los casos posibles, y este constituyente es de los más complejos. La conclusión es de carácter general, y se puede prever que los casos intransitivos serán más sencillos que los transitivos.

```

    si tiene (PTC = (AWAY => ""))
(vi-c1)  => "escapar" + "se"
    si tiene (PTC = (OFF => ""))
(vi-c2)  => "escapar" + "se"
    si tiene (PP = (AFTER => ""))
(vi-d1)  => "seguir"
    si tiene (PP = (FOR => "")) y
    si (NPP de tipo PL)
(vi-d2)  => "dirigir" + "se"
    si no
(vi-e1)  => "contribuir"
    si tiene (PP = (TOWARDS => ""))
(vi-d3)  => "dirigir" + "se"
    si tiene (PTC = (OUT => ""))
(vi-f1)  => "salir" "bien" + flexionar (VERB 0)
    si no
(vi-DEF) => "hacer"
```

Figura 28. Formalización de acepciones (10). MAKE

2.3.3 Conclusiones sobre el ejemplo

Aunando las acepciones transitivas e intransitivas, finalmente obtenemos la descripción de la entrada que aparece en la Figura 29 en la página 75, donde se han incluido las traducciones por defecto para todos los casos.

Con el desarrollo del ejemplo anterior lo que se pretende en primer lugar es mostrar, aunque sea de una manera muy superficial, la diferencia que existe entre la información léxica disponible, en los diccionarios clásicos, y las necesidades que tienen las aplicaciones del PLN, en concreto la traducción automática.

En la formalización no se ha entrado en detalles sobre la notación, ya que lo que se quería era mas bien mostrar el tipo de trabajo de abstracción lexicográfica que ha de llevarse a cabo para la caracterización del proceso de transferencia léxica de las palabras. En el siguiente capítulo se entrará de lleno en esta problemática. Sin embargo, se puede observar que el número de lo que podríamos llamar 'palabras clave' (si, de tipo, es, ...) que hemos utilizado para la descripción es ciertamente limitado, aunque no esté completo. Igual parece suceder con los códigos utilizados (OBJETOP, OBJETO, PP, PTC ...), lo que nos hace pensar en la posibilidad de definir conjuntos de códigos de algún tipo, cerrados, que permitan caracterizar la transferencia léxica de las palabras entre distintos idiomas y ser comprensibles tanto para el usuario como para los programas.

A partir de esta idea se ha desarrollado el formalismo que se presenta en esta memoria, basado en un conjunto de códigos descriptivos relacionados por una sintaxis simple predefinida.

```

    si OBJETO y
    si tiene (PSMODS = NP) y
    si (NP es HUMANO)
(vt-a2)    => "nombrar"
    si no
(vt-a1)    => "hacer" + tr-funcional (OBJETO -> PP "de")
    si tiene (PSMODS = PREDADJ) y
    si (PREDADJ es (READY => ""))
(vt-e7)    => "preparar"
    si no
(vt-e1)    => "hacer"
    si tiene (PSMODS = INFCL)
(vt-i1)    => "obligar" "a" + flexionar (VERB 0) + tr-estructural (INFNP)
    si no
(vt-DEF)   => "hacer"
    si OBJETOP con (OUT => "")
(vt-k2)    => "distinguir"
    si OBJETOP con (UP => "") y
    si es PASIVA
(vt-k4)    => "componer" + NO ACTIVA + ESTÁTICA +
    si tiene (PP con (OF => "por"))
    si no
(vt-k4')   => "constituir"

    si tiene (PTC = (AWAY => ""))
(vi-c1)    => "escapar" + "se"
    si tiene (PTC = (OFF => ""))
(vi-c2)    => "escapar" + "se"
    si tiene (PP = (AFTER => ""))
(vi-d1)    => "seguir"
    si tiene (PP = (FOR => "")) y
    si (NPP de tipo PL)
(vi-d2)    => "dirigir" + "se"
    si no
(vi-e1)    => "contribuir"
    si tiene (PP = (TOWARDS => ""))
(vi-d3)    => "dirigir" + "se"
    si tiene (PTC = (OUT => ""))
(vi-f1)    => "salir" "bien" + flexionar (VERB 0)
    si no
(vi-DEF)   => "hacer"

```

Figura 29. Formalización de acepciones: resultado final. MAKE

Desde el punto de vista más particular de la transferencia léxica, también se pretende poner de manifiesto las dos vertientes que este proceso lleva asociadas:

1. la desambiguación de los diferentes sentidos de las palabras en la lengua origen, y
2. la determinación de las acciones que la transferencia léxica de una palabra origina sobre la estructura de la oración en la lengua meta.

Aunque, por supuesto, no nos hemos enfrentado con todas las posibilidades de transferencia, se han podido detectar algunas condiciones que podemos considerar de carácter general, ya que es evidente que otros verbos, u otro tipo de palabras, se desambiguan en función a éstas. Profundizar en su caracterización es el objeto del capítulo siguiente, pero podemos adelantar en este apartado de reflexión algunas de ellas:

Sobre la LO

En el apartado de las claves sobre la lengua origen hemos detectado que la existencia de objeto directo resulta ser de gran importancia, así como las características de éste: si es preposicional, frasal o simple, si es de un cierto tipo o es una palabra en concreto.

También ha resultado ser de interés alguna característica de la oración en su conjunto, en concreto, en el ejemplo de *"make"* hemos necesitado determinar si la oración en que éste se presentaba era pasiva o no.

En algún momento se ha requerido profundizar en la tipología de los complementos del verbo (si existían postmodificadores o complementos preposicionales), recurriendo a veces, una vez detectados éstos, a averiguaciones del estilo de las descritas en el párrafo anterior: a qué tipo pertenecían, etc.

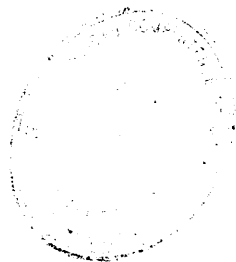
Por lo dicho, podemos imaginarnos otro nivel de desambiguación, que correspondería a toda la serie de interrogantes sobre las características específicas de un constituyente, sea éste del tipo que sea: funcional, como el objeto directo, o sintáctico, como un adjetivo predicativo.

Sobre la LM

En cuanto a los requerimientos que lleva asociados seleccionar una lengua meta en concreto, hemos podido ver que uno de ellos es la determinación de algunos rasgos que tendrán sentido para fases ulteriores a la transferencia léxica propiamente dicha. Entre ellos han surgido el de aquellas traducciones que dan lugar a formar clíticas ("escapar" "se") y el de la traducción selectiva del verbo inglés "to be" por el español "estar".

Ha surgido también un fenómeno de gran importancia como es el de la variación del papel funcional de un constituyente como resultado de la transferencia (ver la oración (2-19) "You should make the previous function a new tool"), de manera que hay que prever mecanismos para que esta transformación se pueda indicar.

Capítulo 3: Un Formalismo de Diccionario Bilingüe



3.1 Representación de la Información

En el capítulo primero de esta memoria se ha considerado la importancia de los problemas relacionados con la representación de la información léxica y de las dificultades que surgen cuando se trata de que una computadora maneje este tipo de datos. En el caso concreto de la traducción automática han de intervenir varios diccionarios en el proceso, destacando por su importancia el que contiene la información bilingüe.

Después de describir en el segundo capítulo el marco del proyecto en el que sitúa el trabajo que se presenta (proyecto MENTOR), y de desarrollar un ejemplo de codificación de la información léxica bilingüe en el que se han puesto de manifiesto algunos de los problemas con los que nos encontramos en el proceso, vamos ahora a describir el formalismo que es el centro de esta memoria, considerando las necesidades lingüísticas que están presentes y los detalles de la solución informática propuesta, que armoniza los requerimientos lexicográficos con los computacionales.

3.1.1 Necesidades Lingüísticas

Las necesidades lingüísticas que contemplamos son de dos tipos. Por una parte, hay que tener en cuenta el punto de vista del lexicógrafo que crea nuevas entradas a partir de la información en lenguaje "natural" y, por otra, el punto de vista de la persona que luego puede tener que actualizar entradas que fueron creadas con anterioridad, o bien corregir errores o completar la información de las mismas a partir de las pruebas de traducción que se realicen. Será obligación del primero exponer su trabajo de forma sistemática y ordenada, de manera que la inspección visual sea inmediata. Por supuesto, cuando la correspondencia entre palabras es uno-a-uno no hay ningún problema; el problema surge cuando la relación no es inyectiva sino suprayectiva, es decir, cuando a una palabra del lenguaje origen le pueden corresponder varias traducciones, como se ha visto en el desarrollo de la extracción de la información léxica del capítulo anterior. ¿De qué depende entonces la elección de una u otra?. En la práctica, la clarificación y determinación de los conceptos léxicos suele requerir el establecimiento de grupos interdisciplinarios compuestos por lingüistas e informáticos, que traten de flexibilizar sus respectivos puntos de vista, y está estrechamente relacionada con los métodos de representación del conocimiento utilizados [Hay85, Aba89]. La creación de grupos de este estilo suele ser una labor muy compleja, puesto que ninguno de sus miembros puede mantenerse al margen de los conocimientos o perspectivas de los demás, ya que esta postura puede dar origen a situaciones de incompreensión absoluta entre unos y otros.

En MENTOR no se pretendía realizar la extracción automática de la información, sino más bien proporcionar al lexicógrafo una herramienta que le permitiera, basándose en criterios puramente lingüísticos, definir las equivalencias múltiples entre las palabras, de manera que estas definiciones pudiesen ser directamente utilizadas por la computadora durante la transferencia léxica. Aunque se partía con la ventaja proporcionada por la restricción del dominio semántico para el que desarro-

llaba el prototipo (que permitía obviar algunos problemas de ambigüedad), sin embargo la solución propuesta debía ser aplicable a cualquier otro dominio, y también extensible para así poder aumentar el rango de las traducciones teóricamente posibles.

De lo que se trataba, por lo tanto, era de realizar un trabajo de investigación sobre las claves de la transferencia léxica entre el par de lenguajes involucrados en nuestro prototipo (inglés como lengua fuente y español como meta), y este trabajo, que atañe principalmente al grupo de lingüistas, debería estar basado sobre una formalización lógica sólida, que garantizase la robustez y completitud del mismo y a la vez se relacionase de forma eficiente con los procedimientos informáticos del sistema, siendo éstos transparentes para el lexicógrafo. El trabajo incluye los niveles más básicos de un sistema de representación y permite que, a partir del nivel formal que se ha desarrollado, se pueda profundizar sobre los aspectos conceptuales del conocimiento [Bra79].

De manera que era necesario que mediante el formalismo propuesto se pudieran armonizar las necesidades procedimentales del sistema con las características más bien de tipo descriptivo utilizadas en las definiciones. Se propuso entonces realizar la descripción de las entradas basando la desambiguación de los distintos sentidos de las palabras en condiciones completamente independientes de la lengua destino y particulares del inglés. Este conjunto de condiciones ha de ser válido para la realización de diccionarios bilingües entre el inglés y otras lenguas destino, puesto que no depende del lenguaje meta, aunque es posible que se requieran nuevas claves de desambiguación de las que no hemos tenido necesidad en la transferencia léxica al español. En cualquier caso, se debería poder determinar un 'superconjunto' de condiciones del que pudiera hacer uso cualquier lengua terminal en un sistema de traducción automática basado como el nuestro en el modelo de transferencia.

En la Figura 30 en la página 81 se describe gráficamente el objetivo de este planteamiento: en el conjunto abstracto de condiciones de desambiguación sobre la lengua origen (LO), que en nuestro caso es el inglés, se utilizan los subconjuntos adecuados en función de la lengua meta a la que se quiera transferir. Estos subconjuntos siguen dependiendo únicamente de la LO.

Sin embargo, el modo de trabajo es realmente el inverso: en vez de intentar determinar todas las condiciones posibles y luego seleccionar las necesarias en cada caso, de lo que se trata es de trabajar desde una de las lenguas meta, perfilando el subconjunto máximo que ésta requiere. Al introducirse una nueva lengua en el sistema, las claves desarrolladas serán parcialmente válidas, detectándose nuevas claves de las que no había habido necesidad en el caso anterior.

También es necesario determinar de alguna manera los efectos que la elección de una u otra traducción provoca en el español como lengua meta. Estas acciones se pueden asimismo precisar, creándose otro conjunto de acciones, que en este caso sólo dependen de la LM.

Los criterios lingüísticos que hemos definido (condiciones sobre la lengua origen) han sido determinados a partir de la información extraída de diferentes diccionarios [Col84, Lar84] y libros de

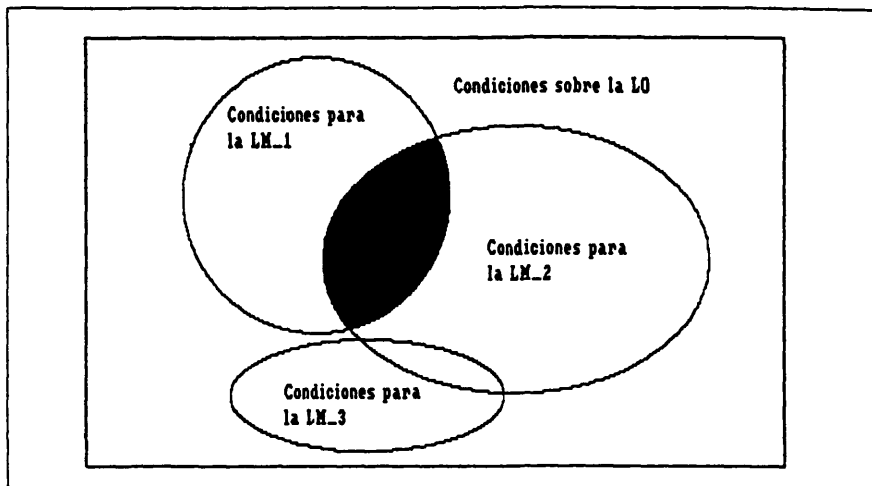


Figura 30. Condiciones sobre la lengua origen.

consulta [Sto65, Pal88, Tho86], entre los que destaca *"A Comprehensive Grammar of the English Language"* [Qui85].

Otro de los requerimientos de este tipo de diccionarios es la necesidad de diseñar interfaces adecuadas con el diccionario, que proporcionen herramientas y mecanismos de acceso y actualización de las entradas ya existentes de una forma cómoda para el lexicógrafo. Lo ideal sería poder disponer de algún tipo de interfaz en lenguaje natural que no obligase al lexicógrafo a realizar una formalización absoluta de las entradas. Aparte de otros muchos, el problema vuelve a ser la materialización de los conceptos. En este sentido, el diseño de la interfaz se plantea de forma similar al diseño de interfaces para sistemas expertos.

3.1.2 Solución computacional propuesta

Teniendo en cuenta las necesidades lingüísticas, se propone un formalismo de diccionario bilingüe que, por una parte, proporciona las claves necesarias para las desambiguaciones de forma concreta y concisa, y por la otra, permite que las entradas expresadas puedan ser así mismo entendidas e interpretadas por la computadora. Este formalismo es además general, válido para distintas lenguas origen y meta, requerimiento cuyas ventajas se justificaban en el apartado anterior.

El formalismo está basado en la descripción de las entradas por medio de elementos mínimos de codificación que se han dado en llamar *Primitivas*, relacionadas entre sí por medio de un conjunto de *Símbolos de Control* que utilizan una sintaxis propia.

Las primitivas, desde el punto de vista del lexicógrafo, entendido éste como la persona que escribe y actualiza las entradas, no son más que una serie de códigos o claves que le permiten expresar de forma semi-descriptiva la información que previamente ha extraído del diccionario o de otras fuentes de consulta. Las hay de dos tipos: *Primitivas de Decisión (PD)*, sólo dependientes de la lengua origen, y *Primitivas de Acción (PA)*, que sólo dependen de la lengua meta. En el primer grupo se incluyen todas aquellas claves necesarias para la discriminación entre los distintos sentidos de la palabra, mientras que en el segundo se encuentran las claves que recogen las acciones que tienen lugar en la palabra (o su entorno) una vez seleccionada una traducción en concreto.

Desde el punto de vista informático, cada primitiva es en sí misma el reflejo de un concepto que tiene sentido en un entorno determinado y que ha de valerse de la información proporcionada por un *analizador en concreto*. Por lo tanto, para su implementación es necesario materializar el concepto asociado a cada una de ellas, relacionando su ejecución con la información de que se dispone en el contexto de la transferencia léxica. Esta materialización muchas veces requiere el reconocimiento de patrones lingüísticos determinados en la estructura o la detección de indicadores en ciertos constituyentes. Por lo tanto, su implementación, en cuanto ésta tiene de estrecha relación con la representación de la información al comienzo de la transferencia léxica, está fuertemente ligada al analizador, a la forma característica en que éste gestiona la información lingüística y, en general, al estilo y los criterios que impone en los análisis. Sin embargo esta estrecha dependencia no influye en la definición de las entradas, ya que los conceptos seguirán siendo válidos aunque cambie la forma concreta de implementación.

Este planteamiento permite independizar las definiciones de los procedimientos que las validan y asegurar que el diccionario, en cuanto es una base de datos, puede ser utilizado y accedido a través de las distintas versiones del analizador o incluso de distintos analizadores. De hecho, uno de nuestros objetivos futuros es probar este mecanismo de representación con otras gramáticas (analizadores de la lengua origen) sin modificar en absoluto la información bilingüe ya creada. Es de esperar que los resultados, por el propio diseño del formalismo, sean altamente satisfactorios.

También es entonces posible adecuar el modo de actuación de las primitivas a las peculiaridades del analizador, de manera que, ejecutadas en un entorno lingüístico adecuado, aunque la gramática cometa algún tipo de errores sistemáticos, éstos puedan ser recuperados mediante la adecuada programación de los procedimientos de las primitivas.

El analizador sobre el que han sido construidas es PEG (PLNLP English Grammar), del que ya se ha hablado en el capítulo anterior, y la representación sobre la que actúan es la estructura arborescente de la que también se ha hablado; todo ello en un entorno de programación LISP, en el que están definidos los procedimientos de cada una de las primitivas, programados, asimismo, como

código LISP. Dicho código se ejecuta al evaluar la definición que se recupera de la base de datos del diccionario durante el período de la transferencia léxica. Esta definición, como se verá en los apartados posteriores, es una estructura marco, adecuada a la sintaxis de los símbolos de control, en la que están presentes las primitivas necesarias.

El resultado de la evaluación es siempre el valor de la traducción del nodo en cuestión, que sistemáticamente se asocia a la variable interna global WORDNODE. De aquí en adelante, WORDNODE será para nosotros sinónimo del nodo a traducir.

A veces se dan casos en los que la traducción de una palabra depende estrechamente de la de otros constituyentes de la oración, es decir, la traducción de una de las palabras determina la de la otra y viceversa. Pueden aparecer entonces traducciones que llamamos "laterales" en el proceso de determinar la de la palabra que originalmente se quería traducir. Para resolver, entre otros, este problema, la solución propuesta consiste en mantener dos variables internas durante todo el proceso:

- **WORDNODE**, asociada siempre al nodo que se está traduciendo, y que es constante durante cada acceso.
- **TESTNODE**, inicialmente asociada también al nodo que se está traduciendo, pero que es variable a lo largo de cada definición. Esta variable está siempre posicionada en aquel constituyente o elemento funcional sobre el que se realizan las discriminaciones en cada caso. Si se producen traducciones laterales durante el proceso siempre afectan al nodo en el que se sitúa TESTNODE, y que llamaremos *foco*.

La finalidad y el sentido de la existencia de estas variables se verá con más claridad en el apartado siguiente.

Por supuesto, ninguna de las primitivas ni de los símbolos de control es necesario que esté presente en las entradas: todos ellos pueden o no aparecer en función de las características específicas del término a traducir.

Tampoco se postula que los conjuntos de primitivas de acción y decisión sean actualmente conjuntos completos, ya que se ha comenzado a trabajar sobre una serie de problemas concretos, a partir de palabras determinadas, principalmente verbos, de forma que se han detectado los conceptos clave asociados a estos. Los subconjuntos definidos cubren sin embargo la práctica totalidad del léxico inglés-español utilizado.

Por último añadiremos que las entradas almacenadas en nuestro diccionario son accedidas mediante una clave doble, compuesta por la propia palabra más la parte de la oración que se desea consultar. Así, consideraremos entradas diferentes a *"like NOUN"* y a *"like VERB"*, o a *"where PRON"* y a *"where CONJ"* (ver el apartado "3.4 Almacenamiento y Acceso" en la página 131 en este mismo capítulo).

3.2 Símbolos de Control

En la lógica del formalismo no sólo es necesario definir las primitivas que hemos mencionado, sino también algunos símbolos de control que doten al mecanismo de flexibilidad y aumenten su potencia. Hemos pretendido limitar el número de éstos al máximo, para no complicar innecesariamente el lenguaje de descripción y conservarlo lo más declarativo posible; sin embargo, hay algunos que ha sido absolutamente necesario introducir para mantener la eficiencia del sistema dentro de unos límites razonables.

Estos símbolos forman parte de la definición de las entradas y, desde este punto de vista, equivalen a conceptos de asociación y relación muy sencillos.

Definición de una entrada:

```
expresión := "traducción" |
  ($)
  + [(expresión_a) | (expresión_b) | (expresión_c)]*
  /* comentario_1 */
  +> (expresión_d)
  /* comentario_2 */

donde:

expresión_a := [ (PD) [(PA)]* ] | expresión
expresión_b := ( & (PD)* [ expresión | expresión_0 | expresión_d ] )
expresión_c := ( /& (PD)* [ (PA) ]* ) expresión_d
expresión_d := [(PA)]* "traducción"
expresión_0 := "traducción" |
  ($)
  [(expresión_a) | (expresión_b)]*
```

Figura 31. Símbolos de control. Sintaxis de definición de las entradas

Los símbolos se pueden considerar de dos tipos: aquellos que tienen significado propio, que son "ejecutables", y aquellos otros cuya misión es simplemente hacer más legibles las entradas del diccionario y no repercuten en absoluto en la evaluación de las mismas; a éstos los llamaremos "no ejecutables".

En la Figura 31 se muestra la sintaxis completa de una estructura de definición, utilizando los símbolos de control que se describen en detalle a continuación. La notación utilizada es la usual, observando que los paréntesis, ')' y '(', no forman parte de la sintaxis sino de las propias expresiones.

- '[' y ']' para indicar parentización corriente y agrupar subexpresiones dentro de una regla de producción
- '~' indica que la cadena afectada puede aparecer ninguna, una o varias veces,
- '!' se utiliza para indicar la 'o' exclusiva de todas las cadenas que relaciona

Se han representado también las marcas no ejecutables (comentarios, etc.), de las que luego hablaremos, en los lugares que habitualmente aparecen, aunque ni su presencia ni la situación que ocupen en la entrada están predeterminadas. "Expresión" se utiliza como símbolo de expresión válida y "traducción" para representar traducción final.

Para definir una entrada del diccionario lo único necesario es realizar la correspondencia entre dicha palabra en el lenguaje origen con una expresión válida descrita con la sintaxis anterior, perteneciendo las PD y PA que se utilicen a los conjuntos de primitivas previamente definidos, y que son conocidos por el lexicógrafo. Estos conjuntos de primitivas se detallan en el apartado siguiente de este mismo capítulo.

El tipo de estructuras que se puede generar es muy rico, dado el carácter recursivo de la definición, desde la más sencilla, en que la traducción es uno a uno, hasta estructuras muy complejas, con un elevado número de recursiones.

A continuación se va a explicar el significado, utilización y sintaxis de los símbolos de control, para lo cual nos apoyaremos en ejemplos ficticios de definición, desde los más sencillos hasta los complejos, introduciendo los mismos problemas que nos fuimos encontrando durante nuestro trabajo en el prototipo, e iremos presentando los símbolos de control que nos permiten resolverlos.

Los ejemplos se han caracterizado con un número y una letra, además de un parámetro que llamamos 'PO', y que corresponde a la *Parte de la Oración*, que en estos ejemplos es genérica. En los casos reales este parámetro será verbo (VERB), nombre (NOUN) o cualquier otro valor posible de los distintos elementos de la oración.

El número del ejemplo indica un caso concreto, y las letras cada una de las posibilidades para formalizar ese caso, manteniéndose las mismas traducciones.

El caso más simple de definición de una entrada es aquel en el que la traducción es directa, "una-a-una". Para esta definición no es necesario utilizar ningún símbolo especial de control, como puede verse en la entrada "ejemplo1A PO".

ejemplo1A PO

"traducción1"

Uno de los primeros problemas que surgen es cómo expresar la situación de dos o más traducciones posibles para la misma palabra. Para ello se creó un símbolo especial de control: \$.

Este símbolo, \$, se utiliza para comenzar cualquier batería de condiciones simples. Su sintaxis es muy sencilla y, aunque hemos visto que no es necesario que aparezca en todos los tipos de entrada, es sin embargo aconsejable como hábito que todas ellas comiencen por él, haya o no condiciones sobre la traducción; de esta forma se simplificarán las actualizaciones cuando se quiera modificar o enriquecer una entrada simple ya existente.

En el "ejemplo1B" se presenta otra forma de definir la entrada anterior haciendo uso del símbolo introducido. Un caso más complejo aparece en "ejemplo2A", donde se ofrecen distintas posibilidades para la traducción de la palabra original.

ejemplo1B PO

```
($  
-> ("traducción1") )
```

ejemplo2A PO

```
($  
+ { (PD1) "traducción1"  
-> { "traducción2" )
```

Las claves para la determinación de una u otra traducción vienen dadas por las primitivas de decisión, PD, que han de aparecer en primer lugar para cada traducción, salvo para la última, que se considera "traducción por defecto". Si las condiciones especificadas no se cumplen en el contexto en el que se va a traducir la palabra, entonces será esta última, la traducción por defecto, la que se dará como válida.

Es aconsejable que siempre exista disponible una traducción por defecto, aunque se tenga la certeza lexicográfica de que nunca se va a dar, porque esta certeza proviene de la consideración de circunstancias y contextos correctos, pero hay que prever que los textos a traducir pudieran no estar perfectamente bien formados: títulos, fracciones de oraciones, etc., como realmente ocurre muy a menudo.

En los dos ejemplos anteriores se pueden observar dos nuevos símbolos: + y +>. Su significado es muy claro: con '+' se señala cada una de las diferentes traducciones, añadiendo '>' cuando a la que se hace referencia es una traducción por defecto de la entrada. Estos símbolos no son ejecutables y su única, y muy importante misión, es facilitar la inspección visual de las entradas, por lo tanto, su uso está recomendado en todas las definiciones.

En cualquiera de los casos es posible indicar una o varias acciones, PA, relacionadas con la lengua meta y provocadas por la elección de una traducción en concreto. Para ello basta incluir las PA necesarias antes de la traducción. Esto, sin embargo, no se puede hacer en el "ejemplo1A", ya que se requiere una estructura mínima de control, y es una razón más para recomendar que las entradas se definan siempre en un marco de condiciones señalado por '\$'. Las nuevas entradas, incluyendo las acciones, son "ejemplo1C" y "ejemplo2B".

ejemplo1C PO

```

($
-> ( (PA1) "traducción1" ) )

```

ejemplo2B PO

```

($
+ ( (PD1) (PA1) (PA2) "traducción1" )
-> ( (PA2) "traducción2" ) )

```

Es posible que existan *condiciones sobre condiciones*, es decir, que una vez que se haya verificado una PD aún no esté completamente desambiguada la entrada y sea necesario tomar en consideración alguna otra condición. Para ello basta con aplicar recursivamente el esquema de condición simple, como puede verse en el "ejemplo3". El resto de las consideraciones siguen siendo válidas a cualquier nivel.

ejemplo3 PO

```

($
( (PD1)
($
+ ((PD2) (PA1) "traducción1" )
-> ("traducción2" ) ) )
-> ( "traducción3" ) )

```

En este último ejemplo podemos ver que existen ya dos traducciones por defecto, correspondientes a los dos niveles de condición que ha sido necesario determinar. Pueden darse muchos más niveles de condición y, también se puede dar el caso de que no todos ellos conduzcan a múltiples traducciones. Es decir, podría ser que, en el ejemplo anterior, la "traducción2" y la "traducción3" fuesen la misma, y que, en realidad lo que queramos expresar no sean dos condiciones simples sino una *condición múltiple*. Para ello se crea un nuevo símbolo: &. La sintaxis de este símbolo es la más compleja y versátil, y se puede utilizar tanto para agrupar varias condiciones de decisión que conducen a una única traducción, como para agrupar condiciones que conducen a varias traducciones.

En el primer caso, la traducción final asignada se puede escribir dentro o fuera del paréntesis de condición múltiple, como puede verse en "ejemplo4A" y "ejemplo4B". En el segundo caso, "ejemplo5A", por propia construcción, al existir una condición simple dentro de la estructura de condición múltiple, las traducciones han de incluirse en el grupo introducido por '&'.

ejemplo4A PO

```
( $
+   ( ( & (PD1) (PD2) "traducción1" )
    ( (PD2)
      ( $
+       ((PD3) "traducción2")
+       ("traducción3" ) )
+>    ( (PA1) "traducción4" ) )
```

ejemplo4B PO

```
( $
+   ( ( & (PD1) (PD2)) "traducción1" )
    ( (PD2)
      ( $
+       ((PD3) "traducción2")
+       ("traducción3" ) )
+>    ( (PA1) "traducción4" ) )
```

ejemplo5A PO

```
( $
    ( ( & (PD1) (PD2)
      ( $
+       ((PD3) "traducción1")
+       ((PD4) "traducción2" ) ) )
+   ( (PD5) "traducción3" )
+>   ( "traducción4" ) )
```

Para dar una idea de la ventaja que supone el uso de condiciones múltiples, en el "ejemplo5B" hemos expresado la misma entrada que en el "ejemplo5A", pero utilizando únicamente condiciones simples. Aquí podemos apreciar cómo es necesario introducir una traducción más, "traducción4", la traducción "por defecto", que en principio no va a suceder nunca, como última posibilidad en todos los casos de condición simple, para poder mantener coherente la estructura. Además hay que considerar en tres lugares diferentes la misma condición, PD5, puesto que ésta es independiente de PD1, PD2, PD3 y PD4 tomadas una a una.

Este ejemplo, aunque es relativamente sencillo, nos da idea de las ventajas de la condición múltiple, sin la cual sería necesario introducir numerosas traducciones innecesarias y, lo que es más importante, obligaría a la definición de entradas sin la más mínima lógica interna.

ejemplo5B PO

```
( $
  ( (PD1)
    ( $
      ( (PD2)
        ( $
          + ( (PD3) "traducción1" )
          + ( (PD4) "traducción2" )
          + ( (PD5) "traducción3" )
          +> ( "traducción4" ) )
        + ( (PD5) "traducción3" )
          +> ( "traducción4" ) )
      + ( (PD5) "traducción3" )
      +> ( "traducción4" ) )
    )
  )
```

Tanto en el Apéndice D, "Algunas entradas completas" en la página 209 como a lo largo de este capítulo y el siguiente se presentan varias entradas reales en cuya definición se ha incorporado el mecanismo de condición múltiple y que ponen de manifiesto claramente sus ventajas. Una de ellas es la correspondiente a "*long*", donde se expresa que si este adjetivo está modificando a un nombre y dicho nombre es "*time*", entonces la traducción obtenida será "*mucho*", que se considera como una excepción del caso general, en el que la traducción es "*largo*" (ver página 217).

Hasta ahora no hemos mencionado la relación que existe entre la sintaxis de los símbolos de control y las variables especiales, WORDNODE y TESTNODE, que se adjudican al nodo a traducir antes de acceder al diccionario. Como ya se dijo, WORDNODE permanece asociado al nodo a traducir mientras dura el acceso (y ejecución) de la entrada correspondiente, y es a este nodo al que se asigna la traducción final obtenida. TESTNODE es, en cambio, variable: partiendo del nodo a traducir, se asocia a los sucesivos *focos de desambiguación* que es necesario recorrer en el proceso de traducción. Por *foco* entendemos aquel elemento de la oración sobre el que recae el peso de la desambiguación, estando siempre relacionado con una primitiva de decisión. Por ejemplo, si se tratara de verificar la transitividad de un verbo (existencia de objeto directo), la primitiva responsable de esta decisión situará el TESTNODE en dicho objeto, en el caso de que exista, pasando a ser el nuevo foco para posteriores consultas.

Tenemos entonces que el cambio de foco está asociado al éxito de las primitivas de decisión, aunque no todas ellas implican una variación del mismo. Un caso de decisión en el que no está involucrado ningún cambio es aquel en el que se consulta sobre alguna característica del propio foco.

Las PD por tanto se encadenan, actuando siempre sobre el TESTNODE actual. De esta forma, cuando se utiliza un marco de condición simple cada nuevo nivel puede trasladar el foco, moviéndonos entre los constituyentes de la oración. En los casos de condición múltiple también se traslada el foco, ya que hemos visto que éstas no resultan ser más que una notación abreviada de las anteriores, pero este traslado ocurre *si y sólo si* todas las condiciones incluidas resultan ciertas.

Las PA, por su parte, pueden actuar tanto en el WORDNODE (la propia palabra a traducir) como en el TESTNODE (la referencia), dependiendo del tipo. Teniendo esto en cuenta, puede ser interesante en algún caso verificar ciertas condiciones sin cambiar el foco de referencia, para que así se determinen las acciones necesarias en los lugares apropiados. Surge así un nuevo símbolo de control: la condición múltiple sin cambio de foco. Ésta se expresa con el símbolo compuesto /&. La sintaxis de este símbolo es similar a la del de condición múltiple con cambio de foco, sólo que ahora no está permitido que las traducciones se expresen dentro del propio paréntesis del símbolo, ya que, si es esto lo que se pretende, lo adecuado es utilizar simplemente &.

ejemplo6A PO

```
($
+ ( (/& (PD1) (PD2) (PA1)) (PA2) "traducción1" )
+ ( (PD3) "traducción2" ) )
-> ( "traducción3" )
```

En el "ejemplo6A", suponiendo que tanto PA1 como PA2 actúan sobre TESTNODE, la primera de ellas tendrá efecto sobre aquel constituyente que sea el foco de referencia después de aplicarse PD1 y PD2, mientras que PA2 actuará sobre el foco de referencia *anterior* a la aplicación de dichas primitivas.

Para terminar, nos quedan únicamente los símbolos /* y */, que se utilizan para delimitar cualquier comentario que se considere de interés incluir en la entrada. Pueden ser ejemplos de uso, anotaciones para el propio lexicógrafo, etc. Los comentarios se pueden añadir en cualquier lugar de la entrada, aunque lo correcto es mantener la definición lo más clara posible. En el "ejemplo6B" se han incluido comentarios para cada traducción, que en este caso serían ejemplos de uso.

ejemplo6B PO

```
($
+ ( (/& (PD1) (PD2) (PA1)) (PA2) "traducción1" )
  /* oración en la que "ejemplo6" se traduce como "traducción1" */
+ ( (PD3) "traducción2" ) )
/* oración en la que "ejemplo6" se traduce como "traducción2" */
-> ( "traducción3" )
  /* oración en la que "ejemplo6" se traduce como "traducción3" */
```


En resumen, disponemos de los siguientes símbolos de control: Símbolos ejecutables

- Condición simple: \$
- Condición múltiple con cambio de foco: &
- Condición múltiple sin cambio de foco: /&

Símbolos no ejecutables

- Marca de traducción condicionada: +
- Marca de traducción por defecto: + >
- Comentarios: /* */

En los ejemplos de las entradas que se presentan a partir de ahora se han omitido los símbolos de condición simple, '\$', pero esta omisión no dificulta la interpretación de las mismas, ya que la presencia de una condición se sigue derivando de los sucesivos sangrados que aparecen en el texto de la definición. Esta decisión se ha tomado para simplificar así la inspección visual de las entradas que se exponen, sin distraer la atención de lo que en cada caso se desea destacar.

3.3 Primitivas

Las primitivas ya hemos dicho que pueden ser de dos tipos: **Primitivas de Decisión** y **Primitivas de Acción**. Las primeras son sólo dependientes del inglés como lengua fuente y las segundas sólo tienen sentido para el español como lengua meta.

En los apartados siguientes pasamos a detallar cada uno de estos grandes grupos: tipo de primitivas en que se subdivide y mecanismo de funcionamiento. Para cada primitiva se proporcionan los parámetros que puede admitir en la llamada, la descripción de su funcionamiento y algunos ejemplos de uso. Estos ejemplos corresponden a porciones de las entradas reales de dichas palabras, que se pueden ver completas en el Apéndice D, "Algunas entradas completas" en la página 209. Asimismo, en el Apéndice A, "Notación" en la página 185, se relacionan los nombres, valores y significado de todos los parámetros utilizados en las llamadas a las primitivas, para su consulta.

En la descripción de los parámetros se ha utilizado una notación estándar, donde se indican, separadas por el símbolo '|', todas las posibilidades. Los paréntesis que se incluyen son necesarios para el funcionamiento correcto de las primitivas, ya que forman parte de la sintaxis requerida por éstas. Entre paréntesis angulares, '<' y '>', se inscriben aquellos parámetros que son opcionales. Como ya se verá, un caso típico de parámetros opcionales es el de los valores de traducción que se pueden adjudicar simultáneamente con la decisión de algunas primitivas. Por ejemplo, si la decisión depende de la presencia de una preposición determinada cuya traducción se conoce, convendrá incluir ésta en la llamada, para evitar recorrer ese nodo una vez más. Una vez se inspeccione el conjunto de primitivas, con sus ejemplos correspondientes, estos mecanismos resultarán más evidentes.

En el caso de las Primitivas de Decisión que, como ya se ha mencionado, pueden producir un cambio en el foco de referencia, se indica explícitamente si éste tiene lugar o no, y, si es así, cuál pasa a ser el nuevo foco.

En cuanto a las Primitivas de Acción, en este caso lo que resulta interesante es conocer sobre qué referencia actúan como resultado de la elección de una traducción determinada: sobre el foco (TESTNODE), la palabra a traducir (WORNODE) u otros elementos. Esta característica también se indica para cada una de ellas.

3.3.1 Decisiones

El fin de las *Primitivas de Decisión (PD)* es determinar si se cumple o no sobre la estructura del análisis la condición a la que está asociado el concepto que materializan; por lo tanto, de su evaluación es de esperar un resultado booleano, que será verdadero ('TRUE' o '1') si es cierta la condición, y falso ('FALSE', 'NIL' o '0') en el caso contrario. Como ya se ha dicho, es habitual que estas condiciones tengan que verificarse sobre elementos de la oración distintos del propio nodo a traducir, WORDNODE, y sea necesario acceder a otros nodos de su entorno, ligados a éste en la estructura de constituyentes o funciones de la oración. Las acciones sobre estos elementos a veces no sólo se limitan a la verificación de condiciones, sino que puede extenderse a la adjudicación de traducciones que vengan dirigidas por la de la palabra que se quiere traducir. Para hacer posibles todas estas acciones, las PD hacen siempre referencia al nodo señalado por TESTNODE, que en un principio coincide con WORDNODE, pero que varía libremente a lo largo de la traducción, visitando los diferentes elementos que intervienen en el proceso de desambiguación. De esta forma es posible asignar cómodamente traducciones a los elementos afectados sin romper el esquema global de definición de la entrada. Por tanto, internamente, otro efecto de la evaluación de las PD es la realización o no del cambio de foco cuando éste tenga sentido, y siempre asociado al valor booleano que proporcionan.

Las condiciones sobre la lengua origen que tienen efecto sobre la traducción pueden ser de varios tipos muy bien diferenciados. Por una parte, puede influir la presencia o ausencia de ciertos elementos funcionales. Uno de los casos más frecuentes es que se obtengan traducciones diferentes si existe o no el objeto directo de la oración, como por ejemplo en el caso del verbo "return".

- (3-1) The thief returned the money.
- (3-2) I returned at ten o'clock.

En (3-1), la traducción ha de ser "devolver", mientras que en (3-2) será "volver", y la elección de una u otra sólo depende de la presencia del objeto directo en la oración.

- (3-1t) El ladrón devolvió el dinero.
- (3-2t) (Yo) volví a las diez en punto.

Este tipo de condiciones, que dependen de la existencia de ciertos elementos funcionales, se han recogido a través de las PD de tipo F (Functional). Pero aún hay otros conjuntos de características que claramente tienen que ver con otras dimensiones de los constituyentes de la oración. Es posible, por ejemplo, que una traducción u otra dependa, no ya de la existencia de los elementos funcionales sino, más concretamente, de ciertas características de los complementos en general.

En (3-3) y (3-4) se puede observar esta nueva dimensión, que es de naturaleza completamente distinta a la anterior.

- (3-3) She became mad when hearing the news.
- (3-4) She became angry when hearing the news.

Aquí se puede ver que la misma construcción puede dar lugar a traducciones completamente diferentes, dependiendo, en este caso, de las palabras concretas que intervienen en la construcción de la oración.

- (3-3t) (Ella) enloqueció al oír las noticias.
- (3-4t) (Ella) se enfadó al oír las noticias.

Esta dimensión del lenguaje se aborda a través de las PD de significado, tipo M (Meaning).

Todavía, sin embargo, se encuentran más dimensiones. Otra de ellas es la relativa a la presencia de ciertos constituyentes, no funcionales, en el entorno de la palabra a traducir. Las PD frasales, tipo P (Phrasal), se encargan de tratar este tipo de condiciones. Ellas suministran las claves necesarias para distinguir entre los casos (3-5) y (3-6), en los que al verbo "be" le corresponden diferentes traducciones.

- (3-5) This machine is two years old.
- (3-6) This machine is the best.

- (3-5t) Esta máquina tiene dos años.
- (3-6t) Esta máquina es la mejor.

Por último, se han detectado también otro tipo de condiciones que dependen de la oración en su conjunto; éstas toman en consideración por medio de las PD de tipo C (Clausal). Las PD/C detectan el aspecto global de la oración: si ésta es negativa o no (característica que no depende únicamente del verbo, sino que puede tener que ver con ciertos pronombres o adverbios), el tiempo en el que se sitúa la acción, etc. Estas características de la oración en su conjunto pueden influir en la traducción de ciertas palabras, como puede verse en (3-7) y (3-8), donde la traducción de la preposición "in" es diferente dependiendo de que la oración sea negativa o no, aunque el resto de los elementos permanezcan iguales.

- (3-7) I didn't see him in years.
- (3-8) I returned in a week.

(3-7t) No lo vi durante años.
(3-8t) Volví tras una semana.

En resumen, las Primitivas de Decisión se articulan en torno a los cuatro ejes de características descritos, y dan lugar a subgrupos semejantes de decisiones de la misma clase. Las primitivas de cada grupo se reconocen por la letra que caracteriza a éste, y que siempre prefija su nombre. Los subgrupos de PD son los siguientes:

- **De Función (Functional).** Estas primitivas tienen en cuenta la existencia y naturaleza de los elementos funcionales de la oración. Hay dos tipos:
 - elementos que aparecen una única vez en la oración (sujeto, distintos tipos de objetos, verbo)
 - aquellos que pueden aparecer más de una vez, asociados a un complemento específico (núcleo de un sintagma preposicional, elemento modificado por un sintagma preposicional)

Hay que resaltar que cuando se habla de elementos funcionales nos referiremos siempre a lo que son los elementos 'profundos', y serán estos los que se detecten a través de las primitivas de tipo F. A modo de ilustración, en los siguientes ejemplos se señala cual será el objeto directo detectado, que en (3-9) coincide con el objeto directo superficial, pero que en (3-10) resulta ser el sujeto, al encontrarse la oración en pasiva.

(3-9) I have just eaten an apple.
(3-10) An apple has just been eaten.

- **De Frase (Phrasal).** Dependen de las características de la complementación de los constituyentes de la oración. Se ocupan de la presencia de ciertos constituyentes (posmodificadores de un sintagma nominal, tipos de complementos funcionales).
- **De Significado (Meaning).** Son primitivas relacionadas directamente con el significado de una palabra, bien por medio de sus características semánticas o morfológicas, bien a través de la palabra misma.
- **De Oración (Clausal).** Tratan aspectos generales de la oración, es decir, factores tales como tiempo, voz o negación, que caracterizan una oración en su conjunto.

Es importante resaltar cómo el foco de referencia cambia al aplicarse las PD. Supongamos que se aplica una PD de tipo F a un nodo verbal. Si el resultado de la evaluación es positivo, entonces la referencia cambia, y la decisión subsiguiente se centrará en el objeto de la evaluación previa (por ejemplo, el sujeto o el objeto directo). De igual manera, algunas PD de tipo P después de una evaluación positiva producen la variación del foco de decisión al objetivo previo (por ejemplo, un complemento adjetivo). Otras PD/P únicamente realizan asignaciones o comprobaciones. En cuanto a las

PD/M y PD/C, éstas no varían el foco de desambiguación, solamente verifican una marca o característica en él o en la oración.

Otra característica importante de la definición de las condiciones es que no existen prioridades adjudicadas entre los diferentes subgrupos, o entre las primitivas de un grupo dado; es decir, los conjuntos y los elementos de cada uno de ellos resultan mutuamente exclusivos (conceptualmente es así, y la implementación realizada lo tiene en cuenta), por lo que, en un mismo nivel de condición, el orden en que se definan los diferentes sentidos sólo depende de las preferencias personales del lexicógrafo o de los criterios estándar predefinidos.

A continuación vamos a describir algunas de las PD más ilustrativas, comentando todos los detalles de funcionamiento y uso en el sistema. En estas descripciones se hace hincapié en el efecto de cambio de referencia, foco, que la actuación de una PD puede provocar.

3.3.1.1 Decisiones de tipo C

C-ASPECT

- **Parámetros:** 'c-aspecto | '(lista-c-aspecto)
- **Descripción:**

En algunos casos resulta necesario especificar el *aspecto*, total o parcial, de la oración, entendiendo por *aspecto* el conjunto de factores gramaticales tales como tiempo, voz, etc., que caracterizan una oración.

En principio, c-aspecto (*característica de aspecto*) puede tomar cualquier valor con sentido; y estos pueden ser de dos tipos diferentes:

- características gramaticales de la oración (presente, negativa, etc.)
- características semánticas de la oración (movimiento, etc.)

En la siguiente tabla se incluyen los posibles valores gramaticales que puede tomar c-aspecto. Cada columna contiene valores mutuamente exclusivos, por lo tanto, el número máximo de parámetros gramaticales en una llamada a C-ASPECT corresponde al número de columnas de la tabla. El conjunto de características semánticas es mucho menos cuantificable y está todavía en estudio, presentándose en el Apéndice B, "Marcas Semánticas" en la página 199 una lista de posibles candidatos.

Al intentar determinar una traducción que puede ser sensible al aspecto de la oración, el conjunto de condiciones puede no ser completo, esto es: quizás el único factor sensible es el tiempo, la negatividad de la oración, etc. De manera que si no se especifica nada sobre alguna característica, el valor de ésta se considera irrelevante.

En resumen, C-ASPECT determina si el aspecto verbal de la oración en la que se encuentra el foco concuerda con el/los c-aspecto/s requeridos.

Figura 32. Aspecto: formas verbales.

Tiempo	Grado1	Grado2	Voz	Negatividad
Pres Past Future	Perfect NOPerfect	Prog NOProg	Passive NOPassive	Neg NONeg

Cambio de foco: No.

• Ejemplos:

anything PRON

- + ((C-ASPECT 'NEG) (G-FEAT 'SGENDER 'NEUTER) "nada")
/* I don't see anything. */
/* (Yo) no veo nada. */

In PREP

- ((F-NOFPP 'NOUN)
- ((M-FEAT 'TME)
- + ((C-ASPECT '(PRES NEG)) "desde hace")
/* I don't see him in years. */
/* No lo veo desde hace años. */
- + ((C-ASPECT '(PAST NEG)) "durante")
/* I didn't see him in years. */
/* (Yo) no lo ví durante años. */
- + ((C-ASPECT '(FUTURE NEG)) "durante")
/* I won't see him in a week. */
/* (Yo) no le veré durante una semana. */
- + ((C-ASPECT '(PAST)) "tras")
/* I returned in a week. */
/* (Yo) volví tras una semana. */
- + ((C-ASPECT '(FUTURE)) "dentro de")
/* I will return in a week. */
/* Volveré dentro de una semana. */
- => ("en"))

like VERB

- + ((F-OBJ)
 - ((C-ASPECT 'PASSIVE) (G-FEAT 'NOPRON) "gustar")
 - /* He has always been liked. */
 - /* Él siempre ha gustado. */

make VERB

- + ((F-OBJ '(up ""))
 - ((C-ASPECT 'PASSIVE)
 - (G-FEAT 'NONACTIVE) (G-STATE) (G-CC 'PP '(of "por")) "componer")
 - /* The group was made up of four people. */
 - /* El grupo estaba compuesto por cuatro personas. */

C-TYPE

- **Parámetros:** 'tipo-oración

- **Descripción:**

Esta primitiva verifica si la oración donde se encuentra el foco es del tipo requerido.

Hasta la fecha, C-TYPE se ha usado principalmente para las entradas de los pronombres.

Cambio de foco: No

- **Ejemplos:**

what PRON

- + ((C-TYPE 'QUES) (G-FEAT 'SNUMBER 'SING) "qué")
 - /* What are you doing? */
 - /* ¿Qué está haciendo? */
- ("lo que")
 - /* Here you have what I wrote yesterday. */
 - /* Aquí tiene lo que escribí ayer. */

whom PRON

- + ((C-TYPE 'QUES) (G-FLEX '(θ)) '("a quién"))
 - /* Whom did you give the book? */
 - /* ¿A quién le dio el libro? */

3.3.1.2 Decisiones de tipo F

F-CC

- **Parámetros:** 'PP < '(prep-ingl "<prep-espa>') > | 'AVP
- **Descripción:**

Esta función identifica el primer complemento circunstancial que aparece al nivel del foco. Es capaz de distinguir entre PPs y AVPs, comportándose de forma ligeramente distinta en ambos casos.

Cuando lo que busca es un PP, si no se ha incluido una preposición determinada en la llamada, F-CC dará un resultado afirmativo para cualquier PP que se encuentre. Si se ha incluido una preposición en concreto, F-CC sólo dará un resultado afirmativo si la preposición que introduce el PP encontrado coincide con la solicitada. Si se ha especificado la traducción de la preposición, "prep-espa" (que puede estar vacía) ésta se asigna simultáneamente. En estos casos se ignoran los posibles AVPs existentes entre el foco y el PP.

Cuando lo que se requiere es un AVP, éste se supone que ha de estar inmediatamente a continuación del foco. Aquí, AVP se utiliza en sentido amplio, e incluye tanto los AVP corrientes como los AVPNP (NP con función adverbial). Por ejemplo "last time" en la expresión: "Last time I did it ...".

Cambio de foco: Sí, al CC.

- **Ejemplos:**

be VERB

```
((P-AJP)
+ ((F-CC 'AVP) "estar")
/* He is very cheerful today. */
/* (Él) está muy contento hoy. */
((F-CC 'PP)
+ ((M-FEAT 'TME) "ser")
/* It will be in the morning. */
/* Será por la mañana. */
```

grow VERB

```
+ ((F-CC 'PP '(outof "en")) (G-FEAT 'CLITIC) "originar")
/* This situation has grown out of a misunderstanding. */
/* Esta situación se ha originado en un malentendido. */
```

make VERB

```
((F-OBJ)
+ ((F-CC 'PP '(into "en")) (G-FEAT 'CLITIC) "convertir")
/* This collection of functions has been made into a new program. */
/* Esta colección de funciones se ha convertido en un nuevo programa. */
((F-CC 'PP '(for "a"))
+ ((M-FEAT 'PL) (G-FEAT 'CLITIC) "dirigir")
/* He made for California. */
/* Él se dirigió a California. */
```

F-INDOBJ

- **Parámetros:** < '(prep-ingl " <prep-espa> ") >
- **Descripción:**

Esta primitiva se usa para posicionar el foco en el Objeto Indirecto de la oración, si éste existe. Una vez allí probablemente se necesitará hacer uso de otras primitivas que verifiquen alguna característica complementaria de éste.

Cambio de foco: SI, al Objeto Indirecto.

- **Ejemplos:**

ask VERB

```
((F-OBJ)
+ ((F-INDOBJ '(from "")) "pedir")
/* She always asks books from me. */
/* Ella siempre me pide libros. */
```

F-IS

- **Parámetros:** función-s | 'PP < '(prep-ingl " <prep-espa> ") >
- **Descripción:**

Esta PD determina si la función sintáctica del foco es la requerida en la llamada. De alguna manera complementa al resto de las decisiones de tipo F.

Son valores admitidos de función-s:

- SUBJ (sujeto)
- OBJ (objeto directo)
- INDOBJ (objeto indirecto)

- VERB (verbo)
- AVP (adverbio)

PP y AVP se admiten tal cual, aunque debieran especificarse: "CC 'PP" y "CC 'AVP". Esta decisión se ha tomado únicamente para no alargar el número de parámetros de F-IS de manera innecesaria.

Esta primitiva se ha revelado muy útil a la hora de traducir preposiciones en función del tipo de núcleo que introducen.

Cambio de foco: No.

• Ejemplos:

— morning NOUN —

- + ((F-IS 'PP '(in "por")) (G-FEAT 'SGENDER 'FEMALE) "mañana")
/* I phoned him in the morning. */
/* (Yo) le telefoneé por la mañana. */

— to PREP —

- ((F-MODPP 'NOUN)
((& (F-MODHEAD 'NOUN)
+ ((F-IS 'PP '(from "desde")) "hasta")
/* DEFWORD eases the shock of moving from the rich REXX */
/* environment to the more austere PL/1 world. */
/* DEFWORD suaviza el impacto de moverse desde el rico
/* entorno de REXX al más austero mundo de PL/1. */

— whom PRON —

- + ((F-IS 'PP) (G-FLEX '(ADJ 0)) '("el" "que"))
/* He is the man to whom you gave the book. */
/* Él es el hombre al que dio el libro. */

— yourself PRON —

- + ((F-IS 'INDOBJ) (G-FEAT 'CLITIC) "se")
/* Give yourself a break. */
/* Dése un respiro. */

F-MODHEAD

- **Parámetros:** 'parte-de-la-oración'
- **Descripción:**

Esta primitiva se usa para posicionar el foco en lo que se supone que es el núcleo modificado por un complemento o elemento de la oración en general, y dará resultado afirmativo si y sólo si éste es del tipo especificado (parte-de-la-oración). Una vez allí se necesitará probablemente consultar otras características de éste.

Cambio de foco: Sí, al núcleo modificado por el complemento.

- **Ejemplos:**

to PREP

```
((F-NOFPP 'NOUN)
  ((& (F-MODHEAD 'NOUN)
    ((M-FEAT 'PROF) "de")
    /* He is assistant to the president. */
    /* Él es ayudante del presidente. */
  + ((F-IS 'PP '(from "desde")) "hasta")
    /* It will last from morning to night. */
    /* Durará desde la mañana hasta la noche. */
  + ((P-PRMODS 'QUANT) "por"))))
    /* They were only ten people to a room. */
    /* Ellos eran solamente diez personas por habitación. */
```

in PREP

```
((F-NOFPP 'NOUN)
+ ((& (F-MODHEAD 'NOUN) (P-PRMODS 'ADJ) (M-MORPH 'SUPR) "de"))
  /* This is the best car in town. */
  /* Éste es el mejor coche de la ciudad. */
+ ((& (P-PRMODS 'NUM) (F-MODHEAD 'NOUN) (M-FEAT 'NUM) "de cada" ))
  /* The current system will benefit two in five people. */
  /* El sistema actual beneficiará a dos de cada cinco personas. */
```

F-NOFADJ

- **Parámetros:** Ninguno
- **Descripción:**

Esta primitiva busca el nombre que gobierna a un adjetivo. Se usa principalmente en la traducción de adjetivos. Como en el caso de otras decisiones F (p.e. F-SUBJ o F-MODHEAD), una

vez realizado el cambio de foco, probablemente se necesitará consultar algunas características complementarias de éste antes de llegar a la traducción definitiva.

Cambio de foco: Sí, al núcleo que gobierna al adjetivo.

• Ejemplos:

```

long ADJ
+ ((& (F-NOFADJ) (M-IS '(TIME "tiempo")) (G-FEAT 'PRENOUN) "mucho"))
/* He was there for a long time. */
/* Él estuvo allí durante mucho tiempo. */
-> ("largo")
/* This road is very long. */
/* Esta carretera es muy larga. */

```

F-NOFPP

• Parámetros: < 'parte-de-la-oración' >

• Descripción:

Esta primitiva busca el núcleo de una frase preposicional. Se usa principalmente para traducir las preposiciones que introducen dichas frases (PP). Si se pasa el parámetro parte-de-la-oración en la llamada, entonces la función sólo dará resultado afirmativo en el caso de que la parte de la oración del núcleo del PP coincida con éste. Al igual que con otras decisiones F, la mera identificación del núcleo no será suficiente en la mayoría de los casos para definir la traducción, y se necesitará consultar otras características del complemento.

Los valores admitidos en este caso para la parte-de-la-oración son NOUN y VERB.

Cambio de foco: Sí, al núcleo del PP.

• Ejemplos:

```

by PREP
+ ((F-NOFPP 'VERB) "")
/* You will get rich by killing your husband. */
/* Se enriquecerá matando a su marido. */

```

In PREP

- + ((F-NOFPP 'VERB) (S-INFNP) "al")
/* In playing football you will get rich. */
/* Al jugar al fútbol se enriquecerá. */
- + ((F-NOFPP 'NOUN)
((M-FEAT 'WTHR) "bajo")
/* I am singing in the rain. */
/* (Yo) estoy cantando bajo la lluvia. */

F-OBJ

- **Parámetros:** < '(p-ingl "<p-esp>") >
- **Descripción:**

Sin parámetros, esta primitiva sólo debe proporcionar un resultado afirmativo en los casos transitivos simples: existe el objeto directo del verbo y el verbo no es transitivo preposicional ni frasal.

Con parámetro, esta primitiva sólo debe proporcionar un resultado afirmativo en las ocurrencias de verbos transitivos y preposicionales o frasales, es decir, existe objeto directo del verbo que está asociado a la p-ingl especificada. Entonces, si se ha proporcionado "p-esp" en la llamada, que puede ser vacía, su valor se asigna directamente.

Como ya se explicó en el apartado dedicado a los diccionarios que utilizamos en MENTOR, para que el objeto directo preposicional se pueda determinar correctamente, p-ingl debe estar incluida en la entrada del verbo en el diccionario inglés monolingüe -dentro del campo OBJTPREP en OD- y para que el objeto directo frasal se determine correctamente, p-ingl debe estar incluida en la entrada del verbo en el diccionario inglés monolingüe -en el campo PTC en OD- (ver página 55).

Cambio de foco: Sí, al objeto de la oración para oraciones activas y al sujeto para pasivas.

- **Ejemplos:**

allow VERB

- + ((F-OBJ)
((P-TYPE 'THATCL) "reconocer")
/* I allow that I made a mistake. */
/* (Yo) reconozco que (yo) cometí un error. */

benefit VERB

- + ((F-OBJ) "beneficiar")
/* This new function benefits the system. */
/* Esta nueva función beneficia al sistema. */

play VERB

- + ((F-OBJ)
((M-IS '(part "cometido")) "desempeñar")
/* She has played a very important part. */
/* Ella ha desempeñado un cometido muy importante. */

want VERB

- + ((F-OBJ)
((P-PSMODS 'INFCL) (S-COMPCL) "querer")
/* I want him to go. */
/* (Yo) quiero que (él) vaya. */

allow VERB

- + ((F-OBJ '(for "")) (G-FLEX '(VERB 0 0)) '("tener" "en" "cuenta"))
/* It will allow for every change. */
/* Tendrá en cuenta todos los cambios. */
- + ((F-OBJ '(of "")) "permitir")
/* It must allow of great hopes. */
/* Debe permitir grandes esperanzas. */

think VERB

- + ((F-OBJ '(of "en")) "pensar")
/* Please, think of it. */
/* Por favor, piense en ello. */

hold VERB

- + ((F-OBJ '(in "")) "contener")
/* This bag holds the money in. */
/* Esta bolsa contiene el dinero. */

look VERB

- + ((F-OBJ '(up **)) "buscar")
/* I usually look some words up in the dictionary. */
/* (Yo) generalmente busco algunas palabras en el diccionario. */

F-SUBJ

- **Parámetros:** Ninguno

- **Descripción:**

Esta primitiva se utiliza para posicionar el foco en el sujeto de la oración. Como en otros casos, una vez allí probablemente será necesario utilizar otras primitivas que consulten alguna característica complementaria de éste.

Hay que hacer notar que se considera que las oraciones pasivas sin complemento agente explícito no tienen sujeto (sujeto profundo).

Cambio de foco: Sí, al sujeto.

- **Ejemplos:**

hold VERB

- ((F-SUBJ))
- + ((M-FEAT 'PHIL) "valer")
/* This condition does not hold. */
/* Esta condición no vale. */
- +> ("resistir"))
/* Wrangler holds if you hold. */
/* Wrangler resiste si usted resiste. */

be VERB

- ((P-AJP)
- ((M-FEAT 'WTHR)
- + ((/(& (F-SUBJ) (M-IS '(it **))) (S-XCHANGE NOUNC) "hacer")
/* It is cold. */
/* Hace frío. */
- + ((/(& (F-SUBJ) ((M-FEAT 'ANIM)) (S-XCHANGE NOUNC) "tener")
/* John is cold. */
/* John tiene frío. */

3.3.1.3 Decisiones de tipo P

P-AJP

- **Parámetros:** Ninguno

- **Descripción:**

Realiza una búsqueda de derecha-a-izquierda en el nivel del foco y a partir de éste. Determina si la primera frase no adverbial es un sintagma adjetivo (*Adjective Phrase*).

Se usa en las entradas de verbos, cuando la presencia de un complemento adjetivo determina, o puede determinar, la traducción de éste.

Cambio de foco: Sí, al AJP.

- **Ejemplos:**

be VERB

```
((P-AJP)
+ ((M-FEAT 'ATTD) (S-XCHANGE NOUNC) "tener")
/* You must be careful when executing this program. */
/* Usted debe tener cuidado al ejecutar este programa. */
```

become VERB

```
((P-AJP)
+ ((M-IS '(mad "")) "enloquecer")
/* She became mad when hearing the news. */
/* Ella enloqueció al oír las noticias. */
```

run VERB

```
((P-AJP)
((F-SUBJ)
+ ((M-FEAT 'CMPT) (G-FEAT 'CLITIC) "ejecutar")
/* These functions will run faster, when you ... */
/* Estas funciones se ejecutarán más rápidamente, cuando usted ...*/
```

P-NP

- **Parámetros:** Ninguno

- **Descripción:**

Su comportamiento es similar al de la primitiva P-AJP, pero identificando un NP.

Cambio de foco: Sí, al NP.

- **Ejemplos:**

be VERB

```
((P-NP)
  ((P-PSMODS 'PREDADJ)
    ((M-IS '(old **)) "tener")
  +   /* This machine is two years old. */
      /* Esta máquina tiene dos años. */
```

make VERB

```
((F-OBJ)
  ((P-NP)
    ((M-FEAT 'HUM) "nombrar")
  +   /* They will make him president. */
      /* Ellos le nombrarán presidente. */
```

P-PRMODS

- **Parámetros:** 'tipo-premod

- **Descripción:**

El objetivo de esta primitiva es determinar si entre los premodificadores del foco existe alguno del tipo indicado. Se usa principalmente para la traducción de nombres, siendo un nombre el foco de partida.

Es capaz de identificar cuatro tipos de premodificadores:

- ADJ (adjetivo)
- NUM (numeral)
- ORD (ordinal)
- QUANT (cuantificador)

Cambio de foco: Sí, al premodificador identificado.

• Ejemplos:

In PREP

- ```
((F-NOFPP 'NOUN)
+ (((& (F-MODHEAD 'NOUN) (P-PRMODS 'ADJ) (M-MORPH 'SUPR) "de"))
/* This is the best car in town. */
/* Éste es el mejor coche de la ciudad. */
+ (((& (P-PRMODS 'NUM) (F-MODHEAD 'NOUN) (M-FEAT 'NUM) "de cada"))
/* The current system will benefit two in five people. */
/* El sistema actual beneficiará a dos de cada cinco personas. */
```

**people NOUN**

- ```
+ ((P-PRMODS 'NUM) "persona")
/* There were two people in the room. */
/* Había dos personas en la habitación. */
```

to PREP

- ```
((F-NOFPP 'NOUN)
+ (((& (F-MODHEAD 'NOUN)
((P-PRMODS 'QUANT) "por"))))
/* There were only ten people to a room. */
/* Había sólo diez personas por habitación. */
```

**be VERB**

- ```
((P-NP)
+ (((& (P-PSMODS 'PP 'IN) (P-PRMODS 'NUM) (G-SETPREP "de") "tener"))
/* The table is 3 metres in length. */
/* La tabla tiene 3 metros de longitud. */
```

P-PRPTCL

• **Parámetros:** Ninguno

• **Descripción:**

Esta primitiva realiza una búsqueda de derecha-a-izquierda al mismo nivel que el foco. Se suele utilizar en las entradas de verbos, cuando la existencia de una oración de participio presente (*Present Participle Clause*) como primer complemento determina la traducción. Se obvian los AVPs que pudiesen existir entre el verbo y la PRPTCL.

Las oraciones de participio presente en inglés (*PRPRTCLs*) no siempre se trasfieren como oraciones del mismo tipo al español; según los casos pueden traducirse como oraciones de infinitivo sustantivadas, oraciones con un verbo en forma personal (en modo indicativo o subjuntivo), oraciones finales o también como oraciones de participio presente. De manera que, aunque durante la transferencia estructural se aplican algunas reglas generales, el caso más frecuente es utilizar alguna Acción S después de esta Decisión P, que determine la transferencia correcta en cada caso. Lo mismo sucede con otras decisiones P.

Cambio de foco: Sí, a la *PRPRTCL*.

• Ejemplos:

keep VERB

```
((F-OBJ '(on ""))
+ ((P-PRPRTCL) "seguir")
/* He keeps on talking. */
/* Él sigue hablando. */
```

P-PSMODS

• **Parámetros:** 'tipo-psmod | '(lista-tipos-psmod) | 'PP 'prep-ingl

• **Descripción:**

Esta primitiva realiza una búsqueda de arriba-abajo, buscando el o los elementos requeridos como postmodificadores en el interior del foco.

Tipo-psmod puede tener los siguientes valores:

- INFCL (oración de infinitivo)
- PRPRTCL (oración de participio presente)
- PREDADJ (adjetivo predicativo)
- PP (complemento preposicional, sin especificar preposición)
- PP 'prep-ingl (complemento preposicional, especificando preposición)
- PTPRTCL (oración de participio pasado)
- RELCL (oración de relativo)

que se identifican de la forma habitual.

Cambio de foco: Sí, al elemento identificado.

• Ejemplos:

ask VERB

- + ((F-OBJ)
 - ((P-PSMODS 'INFCL) (S-COMPCL) "pedir")
 - /* They ask me to write an article. */
 - /* (Ellos) me piden que escriba un artículo. */

anybody PRON

- + ((P-PSMODS '(RELCL INFCL PTPRTCL PRPRCL PP)) "cualquiera")
 - /* Anybody that speaks French can tell you that. */
 - /* Cualquiera que hable francés puede decírselo. */
- + ((P-PSMODS 'PREDADJ) "cualquier")
 - /* Anybody fool can say that. */
 - /* Cualquier loco puede decir eso. */

anything PRON

- + ((P-PSMODS '(RELCL INFCL PTPRTCL PRPRCL PP))
 - (G-FLEX '(8)) "cualquier cosa")
 - /* He will give you anything you ask for. */
 - /* Él le dará cualquier cosa que usted pida. */

be VERB

- ((P-NP)
 - ((P-PSMODS 'PREDADJ)
 - ((M-IS '(old "")) "tener")
 - /* This machine is two years old. */
 - /* Esta máquina tiene dos años. */
 - +> ((S-XCHANGE PREPC "de") "tener"))
 - /* Your string will be two characters long. */
 - /* Su cadena tendrá dos caracteres de longitud. */
 - + ((& (P-PSMODS 'PP 'IN) (P-PRMODS 'NUM) (G-SETPREP "de") "tener"))
 - /* The table is three meters in length. */
 - /* La tabla tiene tres metros de longitud. */

stop VERB

- + ((F-OBJ)
 - ((P-PSMODS 'PRPRICL) (S-COMPCL) "impedir")
 - /* I stopped the book being written. */
 - /* (Yo) impedí que el libro fuese escrito. */
- + ((P-PSMODS 'PP 'FROM)
 - ((P-PSMODS 'PRPRICL) (S-COMPCL) (G-SETPREP "") "impedir")
 - /* I stopped him from writing a program. */
 - /* (Yo) le impedí que escribiese un programa. */

P-PTC

- Parámetros: '(ptc-ingl "<prep-esp>')

• Descripción:

Se usa únicamente para los verbos frasales intransitivos y, como en casos anteriores, "prep-esp" puede ser vacía. Para poder obtener una traducción correcta en estos casos, las posibles partículas del verbo deben estar incluidas en el diccionario inglés monolingüe, en el campo PTC en OD (ver la página 55).

Cambio de foco: No.

• Ejemplos:

be VERB

- + ((P-PTC '(off "")) (G-FEAT 'CLITIC) "ir")
 - /* I am off tomorrow. */
 - /* Me voy mañana. */
- + ((P-PTC '(over "")) (G-FEAT 'CLITIC) "terminar")
 - /* The party is over. */
 - /* La fiesta se ha terminado. */

give VERB

- + ((P-PTC '(up "")) (G-FEAT 'CLITIC) "rendir")
 - /* The enemy gave up after the battle. */
 - /* El enemigo se rindió después de la batalla. */

keep VERB

- + ((P-PTC '(away **))
(G-FEAT 'CLITIC)(G-FLEX '(VERB 0 0))('mantener""a" "distancia"))
/* Keep away from here. */
/* Manténgase a distancia de aquí. */
- + ((P-PTC '(up "atrás")) (G-FEAT 'CLITIC NEG)) "quedar")
/* John is keeping up. */
/* John no se queda atrás. */

make VERB

- + ((P-PTC '(out **)) (G-FLEX '(VERB 0)) ('salir" "bien"))
/* If the demo makes out, the rest will be easy. */
/* Si la demo sale bien, el resto será fácil. */

P-TYPE

- **Parámetros:** 'tipo-foco
- **Descripción:**

Verifica si el tipo del foco corresponde al tipo de oración requerido. Se admiten dos valores: THATCL, oración completiva introducida por "that", y PRPRCTL, notación que, como ya se ha mencionado, se asigna a las oraciones de participio presente que no están introducidas por una preposición.

Cambio de foco: No.

- **Ejemplos:**

allow VERB

- + ((F-OBJ)
((P-TYPE 'THATCL) "reconocer")
/* I allow that I made a mistake. */
/* (Yo) reconozco que (yo) cometí un error. */

keep VERB

- + ((F-OBJ)
((P-TYPE 'PRPRCTL) "seguir")
/* I keep writing five pages every day. */
/* (Yo) sigo escribiendo cinco páginas todos los días. */

3.3.1.4 Decisiones de tipo M

M-FEAT

- **Parámetros:** 'marca-semántica

- **Descripción:**

Verifica la existencia en el foco de la marca semántica requerida. Dichas marcas pertenecen a un conjunto preestablecido y sobre el que actualmente se continúa trabajando. Se pretende determinar familias de marcas relacionadas, para llegar a una definición lo más coherente posible.

En el Apéndice B, "Marcas Semánticas" en la página 199 se proporciona una lista completa de las marcas que estamos utilizando actualmente, junto con una breve descripción.

Cambio de foco: No.

- **Ejemplos:**

allow VERB

```
((F-OBJ)
+ ((M-FEAT 'TME) "dejar")
/* It allows very little time. */
/* Deja muy poco tiempo. */
```

be VERB

```
((P-AJP)
((M-FEAT 'WTHR)
+ (((/8 (F-SUBJ) (M-FEAT 'ANIM)) (S-XCHANGE NOUNC) "tener")
/* Mary is cold. */
/* Mary tiene frío. */
+> ("estar"))
/* The room is cold. */
/* La habitación está fría. */
```



M-IS

- **Parámetros:** '(base-ingl "< base-espa >")

- **Descripción:**

Da resultado afirmativo cuando la base del núcleo del foco actual corresponde a la base-ingl especificada. Si es así, entonces ésta se traduce como "base-espa".

Cambio de foco: No.

- **Ejemplos:**

become VERB

```
((P-AJP)
+ ((M-IS '(mad **)) "enloquecer")
/* She became mad when hearing the news. */
/* Ella enloqueció al oír las noticias. */
```

make VERB

```
((F-OBJ)
((P-PSMODS 'PREDADJ)
+ ((M-IS '(ready **)) "preparar")
/* They always make the program ready. */
/* Ellos siempre preparan el programa. */
=> ("hacer"))
/* She made him happy. */
/* Ella le hizo feliz. */
+ ((M-IS '(mistake "error") "cometer")
/* He made a terrible mistake. */
/* Él cometió un terrible error. */
```

M-MORPH

- **Parámetros:** 'c-morfológica

- **Descripción:**

Determina la existencia en el foco de la marca morfológica requerida en la llamada.

Las marcas morfológicas por las que es posible preguntar son:

- SINGular, PLURal
- MALE, FEMALE, NEUTER
- PERS1, PERS2, PERS3
- DEFInite, INDEFInite

— COMParative, SUPeRlative

Cambio de foco: No.

• Ejemplos:

In PREP

```
((F-MOFPP 'NOUN)
  ((M-IS '(HALF "mitad"))
+   ((M-MORPH 'SING) "por")
/* I discoverd them cut in half. */
/* (Yo) los descubrí cortados por la mitad. */
-> ("en")))
/* I discovered them cut in halves. */
/* Los descubrí cortados en mitades. */
+   (((& (F-MODHEAD 'NOUN) (P-PRMODS 'ADJ) (M-MORPH 'SUPR) "de"))
/* This is the best car in town. */
/* Es el mejor coche de la ciudad. */
```

M-SEG

• Parámetros: 'parte-de-la-oración

• Descripción:

Determina si la parte de la oración del foco corresponde a la especificada.

Se admiten como parámetro todas las posibles partes simples de la oración.

Cambio de foco: No.

• Ejemplos:

he PRON

```
((F-IS 'INDOBJ) (G-FEAT 'CLITIC)
+   ((F-IS 'PP) "él")
/* Give the book to him. */
/* De(le) el libro a él. */
+   (((& (F-OBJ) (M-SEG 'PRON)) "se")
/* Give it to him. */
/* Déselo (a él). */
-> ("le")))
/* Give him the book. */
/* Dele el libro (a él). */
```

3.3.2 Acciones

De la misma manera que los problemas relativos a la precisión de un término pueden ser de varios tipos, dando lugar a las PD/F, PD/P, PD/M y PD/C, los problemas que afectan a la generación, una vez seleccionada la traducción adecuada para la palabra, pueden ser también de varias clases. A modo ilustrativo, examinemos las siguientes oraciones:

- (3-11) They kept at writing this book.
- (3-12) John is keeping up.
- (3-13) She likes him.
- (3-14) I want you to go.

cuyas traducciones son:

- (3-11t) (Ellos) no cesaron de escribir este libro.
- (3-12t) John no se queda atrás.
- (3-13t) A ella le gusta él.
- (3-14t) (Yo) quiero que vayas.

En las dos primeras, (3-11) y (3-12), si observamos la transferencia léxica del verbo, veremos que, para la correcta generación de la oración no basta sólo con determinar la traducción de éste ("cesar" y "quedar atrás" respectivamente), sino que, además, es necesario que las oraciones, que son afirmativas en el original, se conviertan en negativas después de la transferencia. En (3-12), además, el propio verbo tiene un uso pronominal, es decir, debe flexionarse incluyendo el pronombre reflexivo "se". Todos estos aspectos, estrechamente relacionados no ya con la elección de la traducción, sino con la generación de la oración final después de la transferencia léxica, son los que se gestionan a través del primer grupo de Primitivas de Acción: las Primitivas de tipo G (PA/G).

Los dos últimos ejemplos, (3-13) y (3-14), presentan problemas muy diferentes. En (3-13) el verbo se traduce directamente, y no hay que considerar otros factores que tendrían importancia en la generación; sin embargo, como resultado de la elección del verbo español "gustar", tenemos que lo que en inglés resulta ser el sujeto de oración pasa a ser el objeto indirecto y, simultáneamente, el objeto directo pasa a ser el nuevo sujeto. Por lo tanto, en este caso la transferencia léxica provoca un cambio estructural en los elementos funcionales. Del mismo tipo es el problema que se presenta con (3-14): la oración subordinada, que en el original inglés está formada por el infinitivo del verbo, se convierte en español en una oración con verbo personal en subjuntivo; también como consecuencia directa de la transferencia léxica del verbo principal "want".

Para gestionar esta última clase de cambios se creó el segundo grupo de Primitivas de Acción: las Primitivas de Acción de tipo S (PA/S).

En resumen, las *Primitivas de Acción (PA)* se dividen en dos grupos, con las siguientes características:

- **De Estructura (Structural).** Provocan un cambio estructural en la oración resultante (cambio en la parte de la oración de un complemento, relativización de oraciones de infinitivo, etc.)
- **De Generación (Generation).** Añaden ciertas marcas y otro tipo de información general al término traducido, para su utilización en la fase de generación de la oración en la LM (generación del verbo español "estar" en lugar de "ser", generación de pronombres clíticos, etc.)

Como puede observarse, las PA no están directamente relacionadas con el TESTNODE de la manera que lo están las PD. La mayor parte de las PA de tipo G o S lo que hacen es determinar o perfilar alguna característica de generación o estructura, bien del WORDNODE, o bien del entorno o los constituyentes de la oración en su conjunto. Algunas de estas primitivas, sobre todo las de tipo G, proporcionan información que también podría encontrarse en el diccionario de generación, pero que en nuestro sistema se ha elegido incluir via primitivas en el diccionario bilingüe por razones de comodidad. Sin embargo, no excluimos la posibilidad de que en fases posteriores se pueda producir alguna variación, en función de las características y evolución del diccionario de español utilizado.

A continuación se exponen las PA/G y PA/S de la misma forma que en el apartado anterior. Igualmente, los ejemplos de uso que se proporcionan corresponden a porciones de entradas reales que se pueden ver en su totalidad en el Apéndice D, "Algunas entradas completas" en la página 209.

3.3.2.1 Acciones de tipo G

G-CC

- **Parámetros:** 'PP '(prep-ingl "prep-espa") | 'AVP
- **Descripción:**

Esta acción G es formalmente idéntica a F-CC, aunque hasta ahora la opción para adverbios no se ha utilizado.

Busca el primer PP después del foco actual. Si éste está introducido por la preposición prep-ingl, la traducción "prep-espa" le es asignada. En el caso de G-CC, no es optativo proporcionar "prep-espa", porque precisamente su finalidad es traducir la preposición indicada por este valor.

Se usa para algunos PP que pueden aparecer en el contexto de un verbo determinado. Dichos PP no determinan la traducción del verbo, sino que es el propio verbo el que determina la traducción de éstos.

Los PP a los que se refiere G-PP no tienen por qué estar necesariamente presentes en la oración.

Elemento sobre el que actúa: el PP más cercano al foco, si está introducido por prep-ingl.

• Ejemplos:

benefit VERB

```
+> ((G-FEAT 'CLITIC) (G-CC 'PP '(from "de")) "beneficiar")
/* The group benefits from you. */
/* El grupo se beneficia de usted. */
```

move VERB

```
+ ((P-PTC '(away "")) (G-CC 'PP '(from "de")) (G-FEAT 'CLITIC) "alejar")
/* Don't move away from here! */
/* ¡No se aleje de aquí! */
+ ((P-PTC '(in "")) (G-CC 'PP '(to "en")) (G-FEAT 'CLITIC) "instalar")
/* We finally moved in to the new house. */
/* (Nosotros) finalmente nos instalamos en la nueva casa. */
```

G-FEAT

• **Parámetros:** 'c-generación | '(lista-c-generación) | 'c-generación 'genvalue

• **Descripción:**

Introduce algunas marcas características que se usan en la fase de generación. Estas marcas pueden ser de dos tipos, dependiendo de que introduzcan características con uno o varios valores posibles. Si es una característica que sólo puede tomar un valor, entonces basta indicar el nombre de ésta, siendo interpretado el dato de la forma adecuada. Si existen varias posibilidades, es necesario especificar cuál de los valores se atribuye en esa ocurrencia, y es éste el que se adjudicará.

Entre las marcas utilizadas hasta ahora, las que por su naturaleza no requieren que se especifique ningún valor particular son:

- CLITIC, para la generación de pronombres clíticos.
- NEG, para la generación de oraciones negativas.
- NOPRON, para no generar el pronombre "se" en el caso de oraciones pasivas.
- NONACTIVE, para no convertir la oración pasiva a activa (es decir, si la oración es pasiva en inglés, permanece pasiva en español).
- PRENOUN, para aquellos casos en los que el adjetivo debe aparecer delante del nombre.

Los dos siguientes sólo tienen sentido si se definen junto con el valor apropiado.

- SGENDER, utilizado para la determinación del género correcto en el caso de homógrafos castellanos. Por ejemplo: "la mañana" y "el mañana". Valores posibles: MALE, FEMALE y NEUTER.
- SNUMBER, utilizado para forzar el número de una palabra en español, cuando dicho número difiere del que ésta tiene en inglés. Por ejemplo: "people", plural en inglés, cuando se traduce como "gente", singular en español. Valores posibles: SINGular y PLURAl.

Elemento sobre el que actúa: el nodo correspondiente a la entrada a traducir.

• Ejemplos:

keep VERB

- + ((F-OBJ '(at "de")) (G-FEAT 'NEG) "cesar")
/* They kept at writing this book. */
/* Ellos no cesaron de escribir este libro. */
- + ((P-PTC '(up "atrás")) (G-FEAT '(CLITIC NEG)) "quedar")
/* John is keeping up. */
/* John no se queda atrás. */
- + ((F-SUBJ)
(M-FEAT 'PLT) (G-FEAT 'CLITIC) "conservar")
/* These apples keep well. */
/* Estas manzanas se conservan bien. */

like VERB

- + ((F-OBJ)
(C-ASPECT 'PASSIVE) (G-FEAT 'NOPRON) "gustar")
/* He has always been liked. */
/* Él siempre ha gustado. */

make VERB

- + ((F-OBJ '(up ""))
(C-ASPECT 'PASSIVE)
(G-FEAT 'NONACTIVE) (G-STATE) (G-CC 'PP '(of "por")) "componer")
/* The group was made up of four people. */
/* El grupo estaba compuesto por cuatro personas. */

anything PRON

- + ((C-ASPECT 'NEG) (G-FEAT 'SGENDER 'NEUTER) "nada")
/* There wasn't anything near the door. */
/* No había nada cerca de la puerta. */

people NOUN

```
>> ((G-FEAT 'NUMBER 'SING) "gente")
/* There were many people in the room. */
/* Había mucha gente en la habitación. */
```

G-FLEX

- **Parámetros:** '(lista-flex)

- **Descripción:**

Se usa cuando la traducción de una entrada es una expresión multi-palabra. Lista-flex debe contener un elemento para cada palabra de la expresión, y el valor de cada uno de estos elementos debe ser:

- 'O' si la palabra a la que se refiere es una parte invariable de la oración (preposición, conjunción, ...), o si, aunque sea una parte de la oración que generalmente se flexiona (un nombre, por ejemplo) en esa ocurrencia debe aparecer tal cual.
- 'Parte de la Oración' (VERB, NOUN, ADJ) si la palabra a la que se refiere debe ser flexionada dependiendo de las características morfológicas habituales.

Esta acción debe ser la última que aparezca en la definición de la traducción en la que se encuentra (no es necesario, pero sí recomendable). Después de G-FLEX, la traducción debe ser una lista que incluya tantas cadenas de palabras como elementos haya en lista-flex.

Elemento sobre el que actúa: el nodo correspondiente a la entrada que se va a traducir.

- **Ejemplos:**

become VERB

```
((P-NP)
+ ((M-FEAT 'PHOB)
  (G-FEAT 'CLITIC) (G-FLEX '(VERB 0)) '(*convertir" "en"))
/* After a month these flowers will become a wonderful fruit. */
/* Después de un mes estas flores se convertirán en una fruta maravillosa. */
```

grow VERB

```
((P-AJP)
  ((M-IS '(cold ""))
   ((F-SUBJ)
    + ((M-FEAT 'ANIM)
        (G-FLEX '(VERB 0 0 0)) '("empezar" "a" "tener" "frío"))
      /* I am growing cold. */
      /* (Yo) estoy empezando a tener frío. */
```

G-SETPREP

- **Parámetros:** "prep-esp"

- **Descripción:**

Esta acción G debe usarse en combinación con aquellas decisiones relacionadas con la detección de preposiciones o complementos preposicionales (F-CC 'PP, P-PSMODS 'PP, etc.). Es útil en aquellos casos en los que la traducción de una preposición sólo puede ser determinada (o tiene que ser alterada) después de profundizar más en la estructura del PP o en función de otros elementos de la oración.

F-CC 'PP, F-IS 'PP y P-PSMODS 'PP asignan una variable interna a la última preposición considerada, y es sobre esta variable interna sobre la que actúa G-SETPREP.

Elemento sobre el que actúa: la última preposición considerada.

- **Ejemplos:**

keep VERB

```
((F-OBJ)
  ((P-PSMODS 'PP 'FROM)
   + ((P-PSMODS 'PRPRTCL) (S-COMPCL) (G-SETPREP "") "impedir")
     /* The system will keep the user from making mistakes. */
     /* El sistema impedirá que el usuario cometa errores. */
   +> ((S-FCHANGE '(((PREPC INDOBJ))) "ocultar")))
      /* Please, keep these books from Mary. */
      /* Por favor, ocúltele estos libros a Mary. */
```


G-STATE

- **Parámetros:** Ninguno

- **Descripción:**

Se utiliza para determinar el uso del verbo español "estar" en lugar del verbo "ser" en los grupos verbales compuestos del tipo 'BE + pastpart' en inglés. En estos casos, tanto "ser" como "estar" son traducciones posibles, tomándose "ser" como la traducción por defecto.

De lo anterior podemos deducir que G-STATE no tendrá ningún efecto en el caso de encontrarse el verbo en voz 'activa'. Si la voz es 'pasiva', entonces la propiedad 'PASSIVE' en el nodo es sustituida por la propiedad 'STATE'.

Elemento sobre el que actúa: el nodo de la entrada que se va a traducir.

- **Ejemplos:**

```
make VERB
+ ((F-OBJ '(up **))
  ((C-ASPECT 'PASSIVE)
   (G-FEAT 'NONACTIVE) (G-STATE) (G-CC 'PP '(of "por")) "componer" )
  /* The group was made up of four people. */
  /* El grupo estaba compuesto por cuatro personas. */
  ((F-OBJ)
   => ((G-FEAT 'NONACTIVE) (G-STATE) "hacer")))
  /* This table is made of wood. */
  /* Esta tabla está hecha de madera. */
```

3.3.2.2 Acciones de tipo S

S-COMPCL

- **Parámetros:** < 'modo >

- **Descripción:**

Se usa para la transformación de algunas oraciones subordinadas en inglés, compuestas por un NP y un verbo en forma no personal (infinitivo o participio presente) y que dan lugar en español a una oración subordinada cuyo verbo estará en forma personal, siendo el NP el sujeto de la misma. El verbo de la oración principal puede requerir un modo determinado en el de la oración subordinada (indicativo o subjuntivo), que se puede precisar en la llamada.

(3-15) I believe him to be honest.
 (3-15t) Indicativo: (Yo) creo que (él) es honrado.
 (3-16) I want him to go.
 (3-16t) Subjuntivo: (Yo) quiero que (él) vaya.

Si no se proporciona ningún parámetro, se supone que el modo requerido es subjuntivo.

La oración completa se transforma en una oración subordinada sustantiva, introduciéndose en la estructura un nodo THATCOMPL.

Elemento sobre el que actúa: el foco actual, que deberá ser una oración subordinada.

• Ejemplos:

keep VERB

```
((F-OBJ)
  ((P-PSMODS 'PP 'FROM)
    + ((P-PSMODS 'PRPRTCL) (S-COMPL) (G-SETPREP **) "impedir")
      /* The system will keep the user from making mistakes. */
      /* El sistema impedirá que el usuario cometa errores. */
```

want VERB

```
((F-OBJ)
  + ((P-PSMODS 'INFCL) (S-COMPL) "querer")
    /* I want you to go. */
    /* (Yo) quiero que tú vayas. */
```

S-FCHANGE

- Parámetros: '(lista-cambios-función)
- Descripción:

Esta acción S es muy potente y útil para aquellos casos en que la transferencia entre un par de lenguas da lugar a cambios funcionales en los elementos de la oración.

La lista de cambios funcionales puede tener uno o varios elementos. Obviamente, S-FCHANGE sólo debe ejecutarse una vez por traducción.

Los cambios funcionales programados hasta ahora son los siguientes:

- de OBJ (objeto) a PREPC (complemento preposicional): (OBJ PREPC "prep-esp")
- de OBJ (objeto) a SUBJ (sujeto): (OBJ SUBJ)
- de SUBJ a INDOBJ: (SUBJ INDOBJ)

- de PREPC (complemento preposicional) a INDOBJ (objeto indirecto): (PREPC INDOBJ). En este caso, se supone que el foco actual es el complemento preposicional cuya función va a variar.

Elemento sobre el que actúa: todos aquellos elementos funcionales que se indiquen en la llamada.

• Ejemplos:

like VERB

```
((F-OBJ)
-> ((S-FCHANGE '((OBJ SUBJ) (SUBJ INDOBJ)) "gustar"))
/* She likes him. */
/* A ella le gusta él. */
```

keep VERB

```
((F-OBJ)
((P-PSMODS 'PP 'FROM)
-> ((S-FCHANGE '((PREPC INDOBJ)) "ocultar"))
/* Please, keep these books from Mary. */
/* Por favor, ocúltele estos libros a Mary. */
```

play VERB

```
((F-OBJ)
+ ((M-FEAT 'SPRT) (S-FCHANGE '((OBJ PREPC "a")) "jugar")
/* They play football every morning. */
/* Ellos juegan al fútbol todas las mañanas. */
+ ((M-FEAT 'HUM) (S-FCHANGE '((OBJ PREPC "con")) "jugar")
/* They play him at chess. */
/* Ellos juegan con él al ajedrez. */
```

S-INFNP

- Parámetros: Ninguno

- Descripción:

Se usa para forzar la sustantivación de una oración subordinada, cuyo verbo pasa a estar en la forma no personal de infinitivo. Se aplica a oraciones de infinitivo o de participio presente.

Elemento sobre el que actúa: el foco actual, que debe ser una oración subordinada.

- Ejemplos:

make VERB

```
((F-OBJ)
+ ((P-PSMODS 'INFCL) (S-INFNP) (G-FLEX '(VERB 0)) '(*obligar" "a*))
/* The circumstances made him write a new book. */
/* Las circunstancias le obligaron a escribir un nuevo libro. */
```

stop VERB

```
((F-OBJ)
+ ((P-TYPE 'PRPRTCL) (S-INFNP) (G-FLEX '(VERB 0)) '(*dejar" "de"))
/* I had stopped smoking before becoming a doctor. */
/* Yo había dejado de fumar antes de hacerme médico. */
```

S-RELCL

- **Parámetros:** "verb-espa"
- **Descripción:**

En inglés, el gerundio de un verbo puede usarse directamente como adjetivo, pero éste no siempre tiene una traducción directa en español. A veces esta estructura se transforma en una oración subordinada completa. Esta transformación se consigue via S-RELCL. S-RELCL recibe como parámetro un verbo español y convierte el adjetivo, dónde está el foco, en una oración de relativo cuyo verbo es el especificado en la llamada. Por supuesto, dicho verbo debe concordar, en género y número, con el NP referenciado por el adjetivo primitivo.

En realidad, la construcción (S-RELCL "verb-espa") es equivalente a una traducción simple que se adjudicaría al adjetivo en cuestión, y sería sustituible por ésta. Por lo tanto, cuando se utiliza esta primitiva habrá de aparecer en el último lugar de la entrada, en el puesto de la traducción.

Elemento sobre el que actúa: el foco actual.

- **Ejemplos:**

calling ADJ

```
((S-RELCL "llamar"))
/* With REXVAR, the calling exec can ... */
/* Con REXVAR, el exec que llama ... */
```

missing ADJ

```
* ((S-RELCL "faltar"))
/* The police are asking about the missing money. */
/* La policía está preguntando por el dinero que falta. */
```

S-TBASES

- **Parámetros:** '("tbase1" "tbase2")

- **Descripción:**

Se usa para los adverbios, siendo hasta el momento la única acción que permite una traducción múltiple (doble). "tbase1" y "tbase2" son ambas traducciones posibles del adverbio. La elección de una u otra se considera un problema de generación, así que la elección definitiva se pospone hasta esa fase.

Como en el caso de la primitiva anterior, S-RELCL, la finalidad de esta primitiva es proporcionar la traducción de la entrada en la que aparece, lo que la hace equivalente a una traducción simple y obliga a que aparezca en el último lugar de la entrada.

Elemento sobre el que actúa: la entrada a traducir, que debe ser un adverbio.

- **Ejemplos:**

easily ADV

```
* ((G-FEAT 'PSADJ) (S-TBASES '("fácilmente" "con facilidad"))
/* TBASE1: We have easily found errors in the LISP code. */
/* Hemos encontrado fácilmente errores en el código de LISP. */
/* TBASE2: This is an easily calling exec. */
/* Esto es un exec que se llama con facilidad. */
```

S-XCHANGE

- **Parámetros:** 'comp-espa | 'PREPC 'prep-ingl

- **Descripción:**

Se usa para variar la parte de la oración del foco si esto es necesario en el proceso de transferencia léxica (similar a S-FCHANGE).

Las variaciones de la parte de la oración disponibles son las siguientes:

- a PREPC (complemento preposicional): 'PREPC "prep-espa"

- a NOUNC (complemento nominal): 'NOUNC

En los casos programados, se puede observar que es posible transformar un AJP, cuyo núcleo es un adjetivo, en un PREPC o NOUNC, cuyo núcleo es un nombre. Por lo tanto, para poder obtener un resultado correcto es preciso que los adjetivos posean información acerca del nombre con el que están relacionados.

Acerca de 'PREPC, ver S-FCHANGE.

Elemento sobre el que actúa: el foco actual.

• Ejemplos:

be VERB

```
((P-AJP)
+ ((M-FEAT 'ATTD) (S-XCHANGE NOUNC) "tener")
/* You must be careful when executing this program. */
/* Usted debe tener cuidado al ejecutar este programa. */
((P-NP)
((P-PSMODS 'PREDADJ)
+> ((S-XCHANGE PREPC "de") "tener")))
/* Your string will be three characters long. */
/* Su cadena tendrá tres caracteres de longitud. */
```

3.4 Almacenamiento y Acceso

3.4.1 Almacenamiento del Diccionario

El diccionario que hemos descrito en los apartados anteriores se almacena de forma compacta utilizando la metodología DAM, siglas que corresponden a 'A Dictionary Access Method', [Byr86]. Éste es un subsistema de métodos de acceso que proporciona a los programas una forma rápida y adecuada de acceder a grandes archivos, cuya información está asociada a conjuntos de claves. Los diccionarios constituyen un ejemplo típico de este tipo de archivos.

La metodología DAM ofrece una serie de facilidades diseñadas específicamente para las aplicaciones de diccionarios, además de los modos de acceso habituales (secuencial y aleatorio), que incluyen mecanismos de compresión y codificación de la información léxica, y un mecanismo alternativo de acceso a los registros que posibilita la construcción de referencias cruzadas entre diccionarios e indexaciones a los mismos.

La primera versión de DAM se escribió en 1982 y desde entonces ha sido utilizado para muchas aplicaciones, de entre las que se pueden destacar:

- **UDICT:** Es un diccionario de inglés diseñado específicamente para sistemas de procesamiento del lenguaje natural [Hei82], en el que se almacena información sintáctica y semántica de las palabras. Consta de unos 70.000 registros.
- **Diccionarios de Pronunciación:** Almacenan las palabras junto con la parte de la oración que les corresponde así como con su pronunciación, y se han utilizado en dos sistemas de síntesis de voz a partir de texto escrito [Byr85].
- **WordSmith:** Constituye un entorno de acceso a diccionarios en línea, que gestiona diferentes aplicaciones de diccionarios independientes entre sí [Nef88]. Contiene varios diccionarios monolingües y bilingües, que se comprimen y codifican para optimizar el almacenamiento.

En las aplicaciones descritas, las propias entradas constituyen, total o parcialmente, las claves de acceso a la información, dependiendo ésta del tipo de aplicación: definición, pronunciación, datos sintáctico-semánticos, etc.

En nuestro caso, como ya se ha señalado en apartados anteriores, la clave de cada entrada está compuesta por la palabra más la parte de la oración que le corresponde. Esto quiere decir que se mantienen entradas independientes para, por ejemplo, "like" como preposición y "like" como verbo. Todas las palabras comunes están almacenadas en minúsculas, y la parte de la oración en mayúsculas, de acuerdo con las abreviaturas habituales (p.e. PREP para indicar preposición).

La información que se almacena para cada entrada es la descripción que para esa palabra se ha desarrollado de acuerdo con el formalismo descrito anteriormente (compuesto por las primitivas y la sintaxis definida por los símbolos de control).

3.4.2 Acceso

En principio, los mecanismos de acceso a las entradas han de permitir la realización de dos tipos de operaciones:

- **Acceso manual:** enfocado principalmente al lexicógrafo, que le permita examinar las entradas ya existentes, visualizarlas de forma estructurada, modificarlas o introducir otras nuevas, y que, finalmente, le permita también almacenarlas, borrando las versiones antiguas, si las hubiera.
- **Acceso automático:** que permita a los programas acceder directamente a la información de la entrada, considerada como código, recuperándola en una forma fácilmente ejecutable.

El acceso manual ha de ser cómodo y simple de usar, de manera que la tarea de actualizar el diccionario le resulte lo más sencilla posible al lexicógrafo, mientras que el segundo modo de acceso ha de potenciar las características computacionales del sistema.

3.4.2.1 Actualización y Consulta

Para la actualización y consulta del diccionario, que corresponde a un modo de acceso manual, se ha escrito un paquete específico diseñado modularmente, que implementa todas las funciones básicas necesarias para la gestión cómoda de los datos por parte de un lexicógrafo.

Ésta es una interfaz activa, en el sentido de que, en función de los requerimientos del usuario, las opciones que facilita cambian también, adecuándose a las operaciones posibles en cada momento. Al acceder al diccionario a través de ella, puede o no especificarse la parte (o partes) de la oración de la palabra que se desea verificar. Si no se indica ninguna, entonces aparece un menú que exige la fijación de una o varias de éstas.

DIC

DIC palabra <parte-de-la-oración-1> <parte-de-la-oración-2> ...

Figura 33. Acceso al diccionario bilingüe a través de la interfaz

Una vez determinada la palabra y su(s) parte(s) de la oración, se entra en un entorno de edición, en el que, si la entrada ya existía, ésta aparece con el formato habitual (sangrados, etc.), y puede entonces modificarse o borrarse. También se pueden editar entradas nuevas, bien para modi-

ficarias también, o bien simplemente para consultarlas como referencia. Una vez actualizada la entrada, ésta se almacena automáticamente en los archivos correspondientes.

Esta interfaz también permite acceder a los diccionarios de que se disponga en línea, bilingües o monolingües, para poder realizar todo tipo de consultas.

Como ejemplo veamos cómo se puede acceder a una de las palabras del diccionario para consultar su definición, para ello simularemos la consulta de la palabra *"return"*, ejemplo que ha sido tomado de [Zap89].

Si accedemos al diccionario sin especificar la palabra que deseamos consultar, la respuesta que obtenemos es una pantalla en la que se nos requiere dicha palabra, además de la(s) parte(s) de la oración que se desee visualizar:

Introduzca la entrada del diccionario que quiere editar

====>

Seleccione qué parte/s de la oración quiere que se le muestren en su entrada (con cualquier carácter).

nombre
verbo
adjetivo
pronombre
preposición
adverbio
conjunción
datos semánt

Una vez proporcionados los datos requeridos, que también se podían haber dado en la llamada, *'DIC return VERB'*, se nos presenta inmediatamente la definición correspondiente que existe en el diccionario.

Si ésta no existiera, aparecería únicamente la cabecera *return VERB*, y el resto del campo vacío, invitándonos a rellenarlo.

```
return VERB
($
+ ((F-OBJ "devolver")
+ ((F-CC 'PP '(to "a")) "volver")
+> ("volver"))
```

```
1-HELP 2-LINE_ADD 3-QUIT 4-READ 5-UP 1/2 6-FILE 7-ERASE 8-DOWN 1/2
9-ANOTHER ENTRY 11-LINE_DEL 12-WDS_DICT
```

Además de las teclas de función comunes, podemos ver que aparecen otras teclas especialmente programadas para facilitar la labor del lexicógrafo:

- **READ**, que permite leer otra entrada diferente. Esta función es muy útil en aquellos casos en que se está introduciendo una entrada que se sabe es similar a otras ya existentes, con ligeras modificaciones. Por ejemplo, la propia entrada de 'return VERB' puede utilizarse, y, de hecho, es habitual hacerlo, como maqueta para la definición de verbos sencillos, que tienen una única traducción como transitivos y otra como intransitivos.
- **ANOTHER ENTRY**, mediante la cuál es posible editar, independientemente de la entrada actual, otra entrada cualquiera para su modificación o consulta.
- **WDS_DICT**, función a través de la que se pueden acceder otros diccionarios que se tengan disponibles en línea, para ser consultados en el momento de la definición de las entradas. Actualmente se puede consultar mediante esta función el diccionario Collins inglés-español y español-inglés [Col84], y el Longman de inglés [Lon78].

3.4.2.2 Acceso por programa

Desde el programa de transferencia se accede al archivo que contiene el diccionario, que es del tipo DAM, por lo que es necesario que la lectura (única operación necesaria) se realice mediante la ejecución de los mandatos DAM adecuados (abrir el archivo, búsqueda, etc.). Un programa escrito en REXX es el encargado de realizar estas operaciones, recibiendo las instrucciones adecuadas desde el entorno LISP.

Dicho programa recibe el nombre de BIDICT EXEC y, lo que hace en realidad es explorar toda la estructura de diccionarios diseñada en busca de la palabra (palabra + parte de la oración) solicitada, de acuerdo con el orden de prioridades establecido, que es de menos a más específico. Si la palabra se encuentra en alguno de ellos, entonces se deposita la definición que se encuentre asociada a la misma en un archivo, que, una vez devuelto el control al entorno LISP, ACCESS-BIDICT se

encarga de leer y ejecutar, teniendo en cuenta las condiciones (situación en la estructura de la oración, etc.) del nodo al que corresponde la definición de traducción obtenida, para finalizar determinando cuál es su traducción en ese caso. Si no se encuentra la traducción de la palabra solicitada en ninguno de los diccionarios disponibles, entonces BIDICT EXEC, de la misma forma que en el caso anterior, devuelve la propia palabra requerida, la palabra original, pero en mayúsculas. De esta forma se puede detectar fácilmente qué palabras no han sido traducidas en el proceso, y tomar las medidas necesarias.

Una modificación evidente a este modo de operar es proporcionar al usuario del sistema de traducción la posibilidad de incluir las palabras que no se encuentren disponibles en el diccionario en el preciso momento en que éstas sean detectadas durante el proceso de traducción de un texto. Así, permitiendo una interacción más activa del lexicógrafo con el diccionario, los ciclos de prueba se agilizarían en algunos casos. Éste es uno de nuestros próximos objetivos.

Capítulo 4: Aplicación del formalismo en MENTOR

4.1 Ejemplos de representación

Después del capítulo anterior, en el que se ha presentado el formalismo de diccionario bilingüe propuesto junto con el entorno de programación y los mecanismos de implementación desarrollados, en este capítulo se muestra la faceta práctica del mismo, es decir, la forma en que este modo de definición de la información bilingüe se utiliza dentro del prototipo de traducción automática MENTOR.

Aunque en el Apéndice D, "Algunas entradas completas" en la página 209 aparece una selección de las entradas del diccionario, hemos considerado interesante incluir en este punto algunos ejemplos completos. Entre ellos presentamos la formalización del verbo *"make"*, que fue el ejemplo de transferencia léxica para una palabra desarrollado en detalle en el Capítulo 2. Esta formalización se puede comparar con la descripción pseudo-formal que se obtuvo en dicho capítulo, y que aparece en la Figura 29 en la página 75.

4.1.1 RETURN: nombre, adjetivo y verbo

La voz *"return"* es bastante significativa como ejemplo de palabra que puede aparecer actuando con diferentes características sintácticas (adjetivo, nombre y verbo), y cuya traducción es diferente en cada caso; los problemas que se presentan difieren también, y van desde la traducción directa (uno-a-uno) para el nombre, hasta una moderadamente compleja en el caso del verbo, pasando por la perífrasis nominal del caso adjetivo. De hecho, la definición de su traducción como verbo corresponde a un esquema bastante general de traducción, que suele aparecer en otros casos y se utiliza a menudo como referencia: un uso transitivo simple y uno intransitivo, que puede o no ser preposicional.

Como en muchos otros casos, las entradas de adjetivo y verbo pueden definirse de varias maneras que resultarían equivalentes. Aquí hemos representado las que actualmente se encuentran incluidas en nuestro diccionario.

El caso de *"return"* como nombre es bien simple, sólo le hemos adjudicado una traducción, por lo que se ha representado:

return NOUN

"retorno"

Sin embargo, como comentábamos en el capítulo anterior, lo más aconsejable sería definirlo dentro de una condición simple, para facilitar futuras modificaciones, aunque en este caso no parecen muy probables.

El siguiente caso es *"return"* como adjetivo. En esta traducción *"return"* se transforma en un complemento del nombre al que acompaña, por lo que, en nuestro sistema, se le adjudica una traducción perifrástica en la que se indica que no es necesario flexionar la expresión:

return ADJ

+ ((G-FLEX '(0 0)) '("de" "retorno"))

Para finalizar tenemos la acepción como verbo. La principal consideración en este caso es la de la existencia o no de objeto directo del verbo, que nos permite diferenciar entre las traducciones *"devolver"* y *"regresar"* o *"volver"*. Sobre esta diferencia fundamental hemos, además, distinguido las dos traducciones intransitivas que acabamos de mencionar en función de criterios estilísticos, de manera que hemos traducido *"return"* como verbo por *"regresar"* en el caso de que aparezca un complemento circunstancial acompañando a la ocurrencia del verbo (*"return to ..."*) y como *"volver"* en caso contrario. De esta manera obtenemos las siguientes traducciones:

I returned to the city after a long time
(Yo) regresé a la ciudad después de mucho tiempo

en el caso de que se incluya el circunstancial de lugar, y

I returned by car
(Yo) volví en coche

en el caso de que éste no se incluya.

La entrada queda:

return VERB

+ ((F-OBJ) "devolver")
+ ((F-CC 'PP '(to "a")) "regresar")
=> ("volver")

4.1.2 MAKE: verbo

El caso de la transferencia léxica del verbo "make" se ha tratado en detalle en el Capítulo 2 (ver Figura 29 en la página 75), llegando, a partir de la entrada de este verbo en un diccionario, a una definición pseudo-formal de algunas de sus traducciones. A continuación presentamos la entrada que actualmente se incluye en nuestro diccionario bilingüe, en la que es posible discriminar entre 21 sentidos diferentes. Como puede observarse, en esta entrada se han incorporado algunas traducciones posibles que no se consideraron en el ejemplo estudiado en el Capítulo 2, por ser similares a otras cuyas características se analizaron.

make VERB	
(vt-k2) +	((F-OBJ 'out **)) "distinguir"
	((F-OBJ 'up **))
(vt-k4) +	((C-ASPECT 'PASSIVE))
	(G-FEAT 'NONACTIVE) (G-STATE) (G-CC 'PP '(of "por")) "componer"
(vt-k4') =>	("constituir")
	((F-OBJ)
	((P-NP)
(vt-a2) +	((M-FEAT 'HUM) "nombrar")
(vt-a1) =>	((S-FCHANGE '((OBJ PREPC "de")) "hacer"))
	((P-PSMODS 'PREDADJ)
(vt-e7) +	((M-IS '(ready **)) "preparar")
(vt-e1) =>	("hacer")
(vt-i1) +	((P-PSMODS 'INFCL) (S-INFNP) (G-FLEX '(VERB 0)) '("obligar" "a"))
	((M-FEAT 'MONEY) "ganar")
	((M-IS '(mistake "error") "cometer")
	((F-CC 'PP '(into "en")) (G-FEAT 'CLITIC) "convertir")
(vt-DEF) =>	((G-FEAT 'NONACTIVE) (G-STATE) "hacer")
(vi-d1) +	((F-CC 'PP '(after **)) "seguir")
	((F-CC 'PP '(for "a"))
(vi-d2) +	((M-FEAT 'PL) (G-FEAT 'CLITIC) "dirigir")
(vi-e1) =>	("contribuir")
(vi-d3) +	((F-CC 'PP '(towards "a")) (G-FEAT 'CLITIC) "dirigir")
(vi-c1) +	((P-PTC '(away **)) (G-FEAT 'CLITIC) "escapar")
(vi-c2) +	((P-PTC '(off **)) (G-FEAT 'CLITIC) "escapar")
(vi-f1) +	((P-PTC '(out **)) (G-FLEX '(VERB 0)) '("salir" "bien"))
	((P-PTC '(up **)) (G-CC 'PP '(for **)) "compensar")
(vi-DEF) =>	("hacer")

Figura 34. Entrada para el verbo MAKE: acepciones.

A la izquierda de las líneas de la definición se indica a cuáles de los sentidos determinados anteriormente corresponde la descripción que se proporciona.

Podemos establecer cuáles son las similitudes y diferencias entre aquella descripción y la formalizada en la figura anterior, que se derivan inmediatamente de la definición del formalismo descrita en el capítulo precedente:

- En general, las condiciones que en el ejemplo del Capítulo 2 comenzaban por la expresión 'si ...', ahora se indican por la sucesiva indentación de los distintos niveles de paréntesis, que equivalen a condiciones simples.
- La condición sobre el objeto directo simple, que antes se denotaba como OBJETO, ahora se identifica mediante la primitiva (F-OBJ) sin parámetros.
- Los objetos directos preposicionales o frasales (antes OBJETO_P) ahora se detectan mediante (F-OBJ '(p-ingl "<p-esp>*)).
- La existencia de un postmodificador verbal de tipo NP ahora se identifica simplemente mediante (P-NP).
- Para determinar el 'tipo' de un complemento (p.e. si (NP es HUMANO)) se utiliza la primitiva de carácter general M-FEAT, con el parámetro adecuado: (M-FEAT 'HUM).
- Las transferencias estructurales, que en el lenguaje descriptivo del Capítulo 2 se denotaban *tr-estructural*, ahora se indican mediante las primitivas de acción de tipo S. Han aparecido S-INFNP y S-FCHANGE en el ejemplo.
- El resto de los rasgos dependientes de la lengua meta (cómo flexionar una expresión, generar la forma clítica de un verbo, etc...) se indican a través de las primitivas de generación, de las cuales han sido utilizadas en este caso G-FLEX, G-STATE y G-FEAT.

Como puede verse, la transcripción es bastante inmediata, centrándose la complejidad del problema en la descripción de la información léxica en función de un metalenguaje restringido. Una vez realizado este trabajo, que ha supuesto detectar las claves del metalenguaje, y que en nuestro caso se asocia al formalismo de definición (primitivas más símbolos de control), la transcripción o codificación es muy sencilla.

A lo largo de este capítulo se analizan algunos casos de aplicación en los que se utiliza esta entrada: primero en el desarrollo completo de la traducción de una oración en la que se incluye el verbo, y segundo, ejemplificando algunas peculiaridades de uso del formalismo.

4.2 Desarrollo de la traducción de una oración

Una vez expuestas las características tanto del sistema de traducción en su conjunto como del formalismo del diccionario bilingüe diseñado, vamos a tratar en detalle la traducción real de una oración, desde el momento en que se presenta al sistema, en inglés, hasta la obtención del resultado final, es decir, la oración traducida al español.

Se ha seleccionado una oración medianamente compleja, en la que tienen lugar fenómenos típicos de este tipo de proceso y, aunque nos centramos principalmente en lo que es la transferencia léxica, también se resaltarán algunos aspectos relacionados con las demás fases, como son el análisis y la generación.

La oración propuesta es:

"The seven dwarfs allow that Snow-White made them happy for years."

En la Figura 35 se puede ver el árbol de análisis que nos proporciona PEG. El tiempo de CPU invertido en esta operación, incluyendo el acceso al diccionario de inglés, es de 12 segundos.

DECL1	NP1	DETP1	ADJ1*	"the"
		QUANP1	ADJ2*	"seven"
		NOUN1*		"dwarfs"
	VERB1*	"allow"		
	VP1	COMPL1	"that"	
		NP2	NOUN2*	"Snow-White"
		VERB2*	"made"	
		NP3	PRON1*	"them"
		AJP1	ADJ3*	"happy"
	?	?	PP1	PREP1 "for"
				NOUN3* "years"
	PUNC1	"."		

Figura 35. Árbol de análisis proporcionado por PEG. "The seven dwarfs allow that Snow-White made them happy for years."

Para PEG este análisis es sintácticamente correcto, y sólo presenta una ambigüedad en cuanto a qué nivel se refiere el PP "for years". En realidad hay algo que corregir en la oración subordinada, aunque esto quizá es cuestión de criterio. Nos referimos a la forma en que se relaciona "happy" con el resto de la oración. "Make" en esta lectura es un verbo transitivo complejo, que acepta un complemento directo y un adjetivo que actúa a su vez como complemento de este último. Por lo tanto, en

vez de aparecer un único NP, "them happy", complementando al verbo, deberían aparecer dos constituyentes: (NP (PRON)) (AJP (ADJ)) al mismo nivel. De todas formas, como PEG siempre utiliza el mismo criterio en los casos en que aparecen estas estructuras, las primitivas se han diseñado internamente para reconocer este tipo de patrones.

```

SEGTYPE 'SENT'
SEGTYPE2 'DECL'
STR " the seven dwarfs allow that Snow-White made them happy for years"
RULES 4000 4140 5080 7200
RULE 7200 REC VP2 PUNC1
COPYOF VP2 "the seven dwarfs allow that Snow-White made them happy for years"
BASE 'ALLOW'
DICT 'allow'
POS VERB
INDIC PLUR PRES INF INGCOMP NPTOV THATCOMP TRAN PERS3
PRMODS NP1 "the seven dwarfs"
HEAD VERB1 "allow"
PSMODS VP1 "that Snow-White made them happy for years"
PSMODS PUNC1 "."
FRSTV VERB1 "allow"
SUBJECT NP1 "the seven dwarfs"
OBJECT VP1 "that Snow-White made them happy for years"
PREDCOMP VP1 "that Snow-White made them happy for years"
NARGS 1
PARSENO 1
SENTEPE 'DECL'
TOPIC NP1 "the seven dwarfs"
PRED 'ALLOW'
DSUBJECT NP1 "the seven dwarfs"
DOBJECT VP1 "that Snow-White made them happy for years"
NODENAME 'DECL1'

```

Figura 38. Registro DECL1. "The seven dwarfs allow that Snow-White made them happy for years."

Como ya se explicó en el Capítulo 2, el árbol de análisis que PEG muestra no es más que un esquema de la estructura de registros subyacente, en el que se indica la parte de la oración o tipo del constituyente. Nuestro sistema procesa la información contenida en los registros de PEG para conseguir un árbol decorado que constituye una entrada válida para la fase de transferencia. Este proceso se realiza mediante el procedimiento PLTT (ver página 41 en el Capítulo 2), que extrae y codifica la información dependiendo de las características concretas de cada rasgo: si éste es un indicador simple o múltiple, una asociación funcional, etc. Para este ejemplo hemos seleccionado algunos registros significativos de la oración: el de la oración en su conjunto (Figura 36), el de la oración subordinada (Figura 37 en la página 145), el del verbo principal (Figura 38 en la página 146) y el del verbo de la oración subordinada (Figura 39 en la página 146).

Aunque gran parte de la información contenida en los registros es de interés, no todos los rasgos serán necesarios para las fases siguientes de transferencia y generación, por lo que PLTT sólo extrae aquellos que han sido previamente definidos en relación con su utilización posterior. Entre ellos destacamos:

- indicadores de persona y número
- tiempo del verbo y uso (transitivo, intransitivo, ...)
- base y tipo sintáctico del constituyente
- marca de núcleo del nivel (HEAD)
- relaciones funcionales (SUBJECT, OBJECT, ...)

Todos los que se aplican en este caso pueden verse en el árbol decorado correspondiente, Figura 40 en la página 148, única estructura a la que haremos referencia a partir de ahora, y que es el punto de partida de la fase de transferencia. En este árbol se han resaltado en negrita las bases de las palabras de cada uno de los nodos, conjunto sobre el que se realizarán los accesos al diccionario bilingüe.

```

SEGTYPE  'NP'
SEGTYPE2 'VP'
STR       " that Snow-White made them happy for years"
RULES     4888 4858 5888 3395
RULE      3395 COMPL1 VP2
COPYOF    VP3 "that Snow-White made them happy for years"
BASE      'MAKE'
DICT      'made'
POS       VERB ADJ
INDIC     SING PASTPART PAST CAUSTO TRAN PERS3 CLOSED
PRMODS    COMPL1 "that"
PRMODS    NP2 "Snow-White"
HEAD      VERB2 "made"
PSMODS    NP3 "them happy for years"
FRSTV     VERB2 "made"
SUBJECT   NP2 "Snow-White"
OBJECT    NP3 "them happy for years"
ADJ       REC "made"
FOBJ      NP3 "them happy for years"
NARGS     2
COMPZ     COMPL1 "that"
NOMCOMP   VP2 "Snow-White made them happy for years"
ISCOMP    1
PRED      'MAKE'
DSUBJECT  NP2 "Snow-White"
DOBJECT   NP3 "them happy for years"
NODENAME  'VP1'

```

Figura 37. Registro VP1. "The seven dwarfs allow that Snow-White made them happy for years."

```

SEGTYPE 'VERB'
STR      " allow"
COPYOF   REC "allow"
BASE     'ALLOW'
DICT     'allow'
POS      VERB
INDIC    PLUR PRES INF DITRAN INGCOMP NPTOV THATCOMP TRAN
OBJTPREP 'FOR'
OBJTPREP 'OF'
PRED     'ALLOW'
NODENAME 'VERB1'

```

Figura 38. Registro VERB1. "The seven dwarfs allow that Snow-White made them happy for years."

```

SEGTYPE 'VERB'
STR      " made"
COPYOF   REC "made"
BASE     'MAKE'
DICT     'made'
POS      VERB ADJ
INDIC    SING PLUR PASTPART PAST CAUSTO CMPXTRAN DITRAN TRAN
ADJ      REC "made"
OBJTPREP 'AFTER'
OBJTPREP 'FOR'
OBJTPREP 'TOWARD'
OBJTPREP 'TOWARDS'
OBJTPREP 'WITH'
PTC      'OUT'
PTC      'UP'
PRED     'MAKE'
NODENAME 'VERB2'

```

Figura 39. Registro VERB2. "The seven dwarfs allow that Snow-White made them happy for years."

Como ya se expuso, la fase de transferencia se considera que consta de dos partes: transferencia léxica y transferencia estructural. En la Figura 14 en la página 48 se indicaba el esquema del proceso, y se podía ver que ambas están relacionadas, existiendo procesos de transferencia estructural que están provocados o dirigidos por el léxico. Los cambios estructurales que son independientes del léxico son aquellos en los que intervienen esquemas sintácticos típicos. En MENTOR, tal como está diseñada esta fase, estos cambios independientes del léxico se pueden realizar antes, durante o después del acceso al diccionario. Actualmente se está trabajando en el diseño de un algoritmo óptimo para el control de todas las operaciones de transferencia, pero para nuestro ejemplo las consideraremos independientes.

Antes de profundizar en lo que es el acceso al diccionario, recordemos que sobre el árbol se realizan una serie de operaciones previas de transformación, a las que nos referíamos como 'pre-proceso' anterior a la transferencia en el Capítulo II, y que también podríamos llamar 'fase post-PEG', cuyo objetivo es reorganizar algunos aspectos de la estructura del mismo, con la única finalidad de abstraer parte de la información que éste contiene. Entre las operaciones que se realizan está el tratamiento de las oraciones interrogativas y negativas (donde el aspecto negativo se incluye como un atributo más de la oración, eliminándose los nodos superfluos), la unificación de la representación en los casos de coordinación, etc.

El acceso al diccionario se hace a través de un procedimiento LISP llamado ACCESS-BIDICT, al que se le pasan los parámetros adecuados, que han sido previamente extraídos del nodo que se va a traducir. Otro requerimiento importante es que la variable interna WORDNODE haya sido posicionada en dicho nodo, ya que las primitivas harán referencia a ella para las decisiones y acciones que se lleven a cabo. Inicialmente TESTNODE es también WORDNODE, es decir, el nodo con el que se accede.

ACCESS-BIDICT es el procedimiento LISP que accede al diccionario desde el prototipo, y se comunica con un procedimiento externo, escrito en REXX, que recibe la base de la palabra y la parte de la oración con la que ésta aparece, accediendo ya al diccionario compilado. De retorno, ACCESS-BIDICT recibe toda la definición de la palabra, que ejecuta inmediatamente, determinándose la traducción correcta.

El orden en que se acceden los nodos del árbol puede ser importante. En principio, se accede primero el núcleo de cada nivel, en un recorrido de "arriba-abajo", y al resto de los nodos de "izquierda-a-derecha". Pero no sólo los núcleos terminales marcados como tales (*HEAD*) son accedidos, también se acceden algunos nodos padres de un constituyente, aunque no sean terminales. Esto es necesario para determinar algunas relaciones que existen entre constituyentes, como es el caso, p.e., de los objetos directos.

Para realizar un seguimiento del orden en que iremos accediendo los nodos del árbol podemos utilizar tanto la Figura 35 en la página 143, en la que los núcleos de cada nivel (*HEAD*) aparecen marcados con un asterisco, como la Figura 40 en la página 148, en la que éstos aparecen marcados con un par (*ISAHEAD* . 1) En cualquiera de ellas podemos ver que el primer nodo, mirando de arriba-abajo, marcado como núcleo, es el correspondiente al verbo principal de la oración, "allow", y que, como ya hemos visto, recibe el nombre de VERB1.

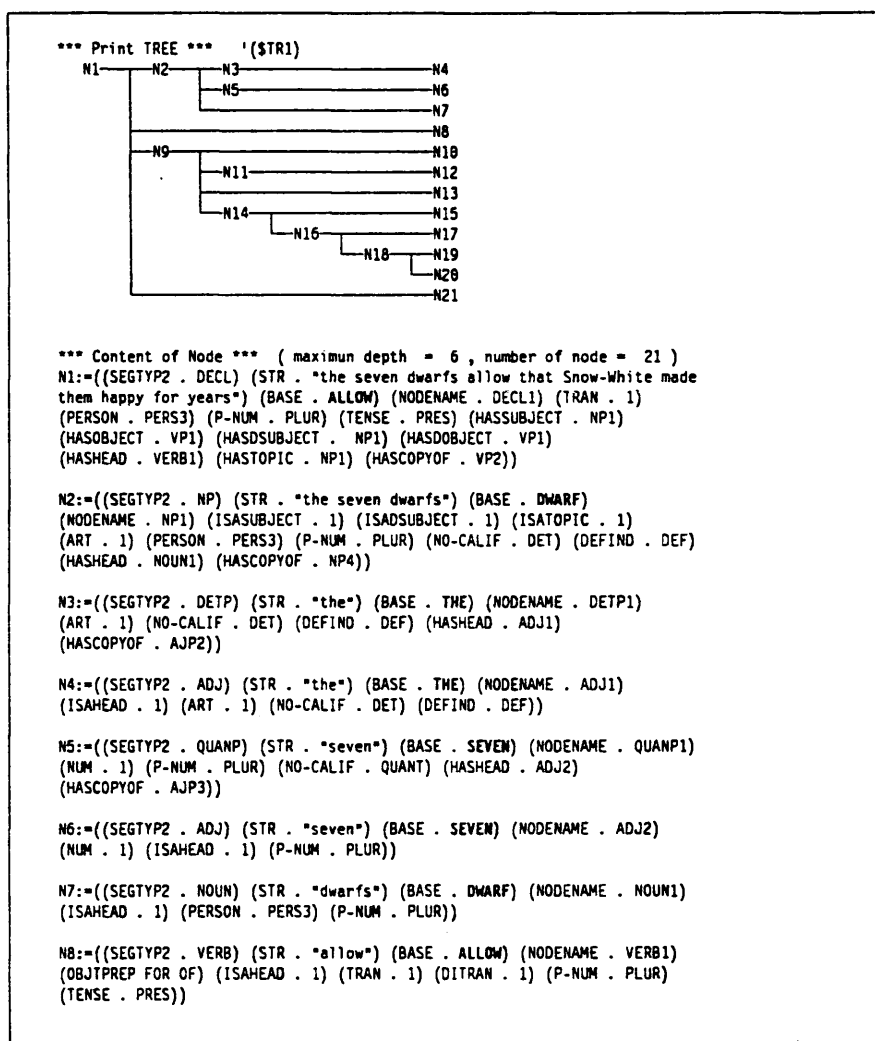


Figura 40. Representación al comienzo de la fase de transferencia. "The seven dwarfs allow that Snow-White made them happy for years."

```

N9:=((SEGTP2 . VP) (STR . "that Snow-White made them happy for years")
(BASE . MAKE) (NODENAME . VP1) (ISAOBJECT . 1) (ISADOBJECT . 1)
(PASTPART . 1) (TRAM . 1) (PERSON . PERS3) (P-NUM . SING) (TENSE . PAST)
(HASSUBJECT . NP2) (HASOBJECT . NP3) (HASDSUBJECT . NP2)
(HASDOBJECT . NP3) (HASHEAD . VERB2) (HASCOPYOF . VP3))

N10:=((SEGTP2 . COMPL) (STR . "that") (BASE . THAT) (NODENAME . COMPL1))

N11:=((SEGTP2 . NP) (STR . "Snow-White") (BASE . SNOW-WHITE)
(NODENAME . NP2) (ISASUBJECT . 1) (ISADSUBJECT . 1) (PRPRNOUN . 1)
(PERSON . PERS3) (P-NUM . SING) (HASHEAD . NOUN2) (HASCOPYOF . NP5))

N12:=((SEGTP2 . NOUN) (STR . "Snow-White") (BASE . SNOW-WHITE)
(NODENAME . NOUN2) (ISAHEAD . 1) (PRPRNOUN . 1) (PERSON . PERS3)
(P-NUM . SING) (HASCOPYOF . NOUN4))

N13:=((SEGTP2 . VERB) (STR . "made") (BASE . MAKE) (NODENAME . VERB2)
(OBJTPREP AFTER FOR TOWARD TOWARDS WITH) (PTC OUT UP) (ISAHEAD . 1)
(PASTPART . 1) (TRAN . 1) (DITRAN . 1) (TENSE . PAST))

N14:=((SEGTP2 . NP) (STR . "them happy for years") (BASE . THEY)
(NODENAME . NP3) (ISAOBJECT . 1) (ISADOBJECT . 1) (ANIM . 1)
(PERSON . PERS3) (P-NUM . PLUR) (HASHEAD . PRON1) (HASPREDADJ . AJP1)
(HASCOPYOF . NP6))

N15:=((SEGTP2 . PRON) (STR . "them") (BASE . THEY) (NODENAME . PRON1)
(ISAHEAD . 1) (ANIM . 1) (PERSON . PERS3) (P-NUM . PLUR))

N16:=((SEGTP2 . AJP) (STR . "happy for years") (BASE . HAPPY)
(NODENAME . AJP1) (ISAPREDADJ . 1) (HASHEAD . ADJ3) (HASCOPYOF . AJP4))

N17:=((SEGTP2 . ADJ) (STR . "happy") (BASE . HAPPY) (NODENAME . ADJ3)
(ISAHEAD . 1))

N18:=((SEGTP2 . PP) (STR . "for years") (BASE . YEAR) (NODENAME . PP1)
(TME . 1) (PERSON . PERS3) (P-NUM . PLUR) (HASHEAD . NOUN3)
(HASPPOBJ . NP7) (HASCOPYOF . NP7))

N19:=((SEGTP2 . PREP) (STR . "for") (BASE . FOR) (NODENAME . PREP1))

N20:=((SEGTP2 . NOUN) (STR . "years") (BASE . YEAR) (NODENAME . NOUN3)
(ISAHEAD . 1) (TME . 1) (PERSON . PERS3) (P-NUM . PLUR))

N21:=((SEGTP2 . PUNC) (STR . ".") (BASE . .) (NODENAME . PUNC1)
(HASCOPYOF . PUNC2))

```

Figura 41. Representación al comienzo de la fase de transferencia (cont.). "The seven dwarfs allow that Snow-White made them happy for years."

En nuestro diccionario, la entrada de "allow", una vez eliminados los símbolos de control, resulta ser:

allow VERB

```
+ ((F-OBJ '(for "")) (G-FLEX '(VERB 0 0)) ('"tener" "en" "cuenta"))
+ ((F-OBJ '(of "")) "permitir")
((F-OBJ)
+ ((P-TYPE 'THATCL) "reconocer")
+ ((M-FEAT 'TME) "dejar")
+ ((P-PSMODS 'INFCL) (S-COMPCL) "permitir")
-> ("permitir")
-> ("permitir")
```

La primera condición que se comprueba es la existencia de un objeto directo preposicional introducido por la preposición "for" y, posteriormente, por la preposición "of". No es éste el caso aquí, puesto que, aunque el verbo es transitivo, no es sin embargo preposicional. Se pasa entonces a evaluar la condición sobre la existencia de un objeto directo simple. Esta vez el resultado es positivo, como puede verse fácilmente en la Figura 40 en la página 148, donde el nodo N9, que corresponde a la oración completiva, está marcado con (ISAOBJECT . 1) en su nivel. El efecto inmediato es el cambio de TESTNODE, que pasa a situarse en lo que es el objeto: *"that Snow-White made them happy for years"*. Las siguientes condiciones, si las hay, se verificarán por lo tanto sobre este constituyente.

Efectivamente, hay otro grupo de condiciones que han verificarse antes de que se pueda determinar la traducción, la primera de las cuales tiene que ver con las características sintácticas del constituyente. Exactamente, lo que se requiere es que éste sea de tipo THATCL, es decir, una oración completiva (*"that"* como nexos). Esta condición se verifica mediante la primitiva P-TYPE con el parámetro THATCL. Como éste es el caso, P-TYPE devuelve un valor afirmativo y, a continuación nos encontramos ya la traducción seleccionada, que resulta ser el valor de la lista de condiciones que comenzaba con F-OBJ, y por lo tanto, al ser un valor no nulo, también será el valor de la condición global y, en consecuencia, la traducción de "allow" para esa ocurrencia: *"reconocer"*.

La traducción que acabamos de conseguir, y todas las siguientes, se incluye en el árbol decorado como una nueva propiedad, cuyo nombre genérico es TBASE (de Translated BASE). Así que, en el nodo correspondiente a "allow", N8, tenemos ahora un nuevo rasgo (TBASE . "reconocer"), como se puede observar en la representación parcial de la Figura 42.

```
N8:=((SEGTYPE2 . VERB) (STR . "allow") (BASE . ALLOW) (NODENAME . VERB1)
(OBJTPREP FOR OF) (ISAHEAD . 1) (TRAN . 1) (DITRAN . 1) (P-NUM . PLUR)
(TENSE . PRES) (TBASE . "reconocer"))
```

Figura 42. Asignación de la traducción de "allow". "The seven dwarfs allow that Snow-White made them happy for years."

Falta comentar que el sistema mantiene una tabla hash con los resultados (sin evaluar) de todos los accesos al diccionario. Esta tabla se consulta en ACCESS-BIDICT, antes de comunicarse con los procedimientos externos, de forma que si una palabra ya ha sido consultada alguna vez durante la sesión no es necesario repetir el proceso, leyéndose su definición directamente de la tabla. Cada lectura de la misma definición puede dar lugar a traducciones diferentes, ya que las condiciones varían en cada caso (WORDNODE, TESTNODE y el propio árbol pueden ser distintos).

El siguiente constituyente que se traduce es NP1: "*The seven dwarfs*", comenzando con "*dwarfs*", que es el núcleo. El acceso se realiza con la base de la palabra, que PEG ha sabido identificar correctamente como "*dwarf*". En este caso la traducción es uno-a-uno, y se ha determinado que sea "*enanito*", "*The*" y "*seven*" también se traducen directamente.

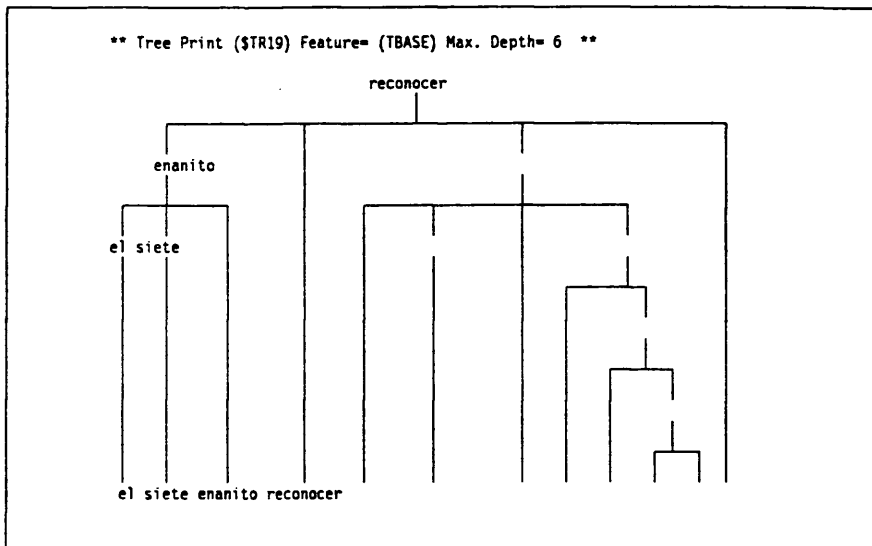


Figura 43. Transferencia Léxica (1). "The seven dwarfs allow that Snow-White made them happy for years."

El árbol que tenemos traducido en estos momentos aparece en la Figura 43, donde se puede apreciar cómo la traducción se propaga a otros nodos que dominan a los traducidos. Los nodos que aún no han recibido traducción permanecen vacíos.

Siguiendo el recorrido de izquierda-a-derecha por los núcleos de cada nivel, el siguiente nodo que se traduce es el correspondiente a "made", N13 en la Figura 40 en la página 148. Desde el sistema se accede al diccionario de la forma que ya se ha descrito, con la base correspondiente, que

PEG identifica correctamente como "make". La definición de este verbo que se encuentra es la que se representó en el apartado anterior, y que recordamos era:

make VERB	
+	((F-OBJ 'out **)) "distinguir"
	((F-OBJ 'up **))
+	((C-ASPECT 'PASSIVE))
	(G-FEAT 'NONACTIVE) (G-STATE) (G-CC 'PP '(of "por")) "componer"
→	("constituir")
	((F-OBJ
	((P-NP
+	((M-FEAT 'HUM) "nombrar")
→	((S-FCHANGE '((OBJ PREPC "de"))) "hacer")
	((P-PSMODS 'PREDADJ
+	((M-IS 'ready **)) "preparar"
→	("hacer")
+	((P-PSMODS 'INFCL) (S-INFNP) (G-FLEX '(VERB 0)) '("obligar" "a"))
+	((M-FEAT 'MONEY) "ganar")
+	((M-IS 'mistake "error") "cometer")
+	((F-CC 'PP '(into "en")) (G-FEAT 'CLITIC) "convertir")
→	((G-FEAT 'NONACTIVE) (G-STATE) "hacer")
+	((F-CC 'PP '(after **)) "seguir")
	((F-CC 'PP '(for "a"))
+	((M-FEAT 'PL) (G-FEAT 'CLITIC) "dirigir")
→	("contribuir")
+	((F-CC 'PP '(towards "a")) (G-FEAT 'CLITIC) "dirigir")
+	((P-PTC 'away **) (G-FEAT 'CLITIC) "escapar")
+	((P-PTC 'off **) (G-FEAT 'CLITIC) "escapar")
+	((P-PTC 'out **) (G-FLEX '(VERB 0)) '("salir" "bien"))
+	((P-PTC 'up **) (G-CC 'PP '(for **)) "compensar")
→	("hacer")

En el momento de acceder, tanto TESTNODE como WORDNODE están situados en N13. Las dos primeras decisiones, que verifican la existencia de objetos directos preposicionales, dan resultado negativo, de manera que llegamos a la tercera, F-OBJ simple. El resultado de ésta es afirmativo, y TESTNODE pasa a situarse en lo que está identificado como el objeto del verbo que, como puede verse en la Figura 37 en la página 145 y en la Figura 40 en la página 148, es el NP "them happy for years", que corresponde al nodo N14.

Este análisis admite variaciones, pudiendo ser que lo único que se identificase como objeto fuese "them". En cualquier caso aunque el análisis variase la entrada del verbo seguiría siendo válida, por construcción de nuestro formalismo.

Ahora, con TESTNODE en el objeto, se plantea otra batería de condiciones que es necesario resolver para llegar a la traducción. La primera está representada por la primitiva P-NP. Las decisiones de tipo P cuestionan la presencia en la oración de determinados constituyentes en el plano

sintáctico, no en el funcional. En concreto, P-NP dará un resultado afirmativo si el siguiente constituyente por la derecha del TESTNODE actual, obviando adverbios, es un NP cualquiera. Si fuera el caso, TESTNODE pasaría a posicionarse en éste. El resultado es sin embargo negativo, así que se pasa a evaluar la siguiente condición, que está representada por P-PSMODS.

P-PSMODS, que decide sobre los PoStMODificadores del TESTNODE, requiere un parámetro que indica el tipo de éstos que ha de decidirse. En este caso el parámetro es PREDADJ, abreviatura de PREDicative ADJective (Adjetivo Predicativo).

Se puede pensar que P-NP y P-PSMODS son semejantes desde el punto de vista lógico, ya que ambas deciden sobre la existencia de postmodificadores del TESTNODE actual. Efectivamente, quizá convenga unificarlas en una única primitiva parametrizada, al estilo de la P-PSMODS actual, y no hay ninguna razón para no hacerlo. Históricamente P-NP es anterior, y cuando se creó todavía no nos habíamos enfrentado con los casos que luego se resuelven con P-PSMODS. Estos casos (ver la lista de parámetros posibles en el Apéndice A, "Notación" en la página 195) corresponden siempre a un tipo de constituyentes cuyo análisis resulta equivoco o ambiguo para PEG, que sistemáticamente los incluye como complementos del constituyente que les precede. Por esta razón el análisis difiere (desde el punto de vista de PEG) del que sería deseable. P-PSMODS tiene en cuenta esta situación y, puesto que es un hecho sistemático, se puede abordar realizando la búsqueda del postmodificador requerido entre los postmodificadores hijos del TESTNODE en vez de entre los hermanos. Según PEG vaya analizando estos casos con mayor precisión, P-PSMODS podrá ir evolucionando, sin necesidad de modificar las entradas que ya estén realizadas, sólo cambiando la ejecución de la primitiva. En cualquier caso, la unificación de P-NP con ésta no ofrece ningún problema, y se han mantenido separadas únicamente por comodidad.

Volviendo al ejemplo, y teniendo en cuenta lo dicho anteriormente, es evidente que se reconoce un adjetivo predicativo entre los postmodificadores del objeto: el hijo *"happy for years"* de N14, que, como puede verse en la Figura 41 en la página 149 aparece marcado como (ISAPREDADJ . 1) y corresponde al nodo N16. Por lo tanto el resultado es afirmativo y se pasa a evaluar la siguiente batería de condiciones, posicionándose de nuevo TESTNODE.

El que aparezca un PP, *"for years"*, como parte del adjetivo predicativo, aún siendo incorrecto, es una situación que se tiene en cuenta y se puede considerar, ya que este problema sobre la ambigüedad de los PPs a la hora del análisis (¿qué nivel les corresponde?) es general y la solución de PEG también es general, asociándolos siempre al último nivel analizado.

Ahora se realiza una discriminación acerca del tipo de adjetivo presente. Si éste es *"ready"*, como por ejemplo en: *"they can make the program ready"*, la traducción tiene que ser: *"(ellos) pueden preparar el programa"*, eliminándose el adjetivo, que no aparece en la traducción española. En los demás casos estudiados (habrá más, aún no detectados) el verbo *"make"* se acaba traduciendo por el español *"hacer"* y el adjetivo permanece también en la estructura.

M-IS por lo tanto verifica la base del nodo en el que se encuentra TESTNODE. Dicha base es HAPPY y no READY, así que se pasa a evaluar la siguiente condición que resulta ser ya la traducción por defecto en este caso: "hacer". Así que éste es el valor de la condición asociada a los postmodificadores y, por lo tanto, de la asociada al objeto simple y también de la definición total del verbo.

Como resultado de esta operación se traduce la base de N9 y N13.

Los siguientes nodos en el proceso son N10 y N12: "that" y "Snow-White", cuya traducción es directa, siendo sus TBASEs respectivas: "que" y "Blancanieves". Pasamos ahora al subárbol situado a la derecha del verbo.

En elemento más a la izquierda de este subárbol es el pronombre personal "them", cuya base es la forma principal del mismo: "they". Esta palabra nos da pie a comentar el tratamiento que se hace de los pronombres en el sistema. Aunque es discutible, parece ser que lo más puro sería traducir los pronombres personales durante la transferencia base-por-base, es decir, "they" por "ellos", dejando la flexión del pronombre para la fase de generación. Hoy por hoy sin embargo esto se hace durante la fase de acceso al diccionario bilingüe por razones de comodidad y modularidad, ya que el diccionario de que disponemos para la generación no tiene unificados los pronombres personales bajo una forma base única, y tendría que realizar esta flexión un módulo ad hoc. Puestos en este caso, nos resulta más sencillo utilizar los mecanismos de definición de transferencias que nos proporciona el formalismo del diccionario y, teniendo perfectamente definido el lugar donde se realiza esta operación, poder eliminar la información superflua, simplificando las entradas, cuando ya no sea necesario. De todas formas, este principio no es general para todos los pronombres, baste comparar las entradas de "they" y, por ejemplo, "anybody":

they PRON

```
+ ((F-IS 'SUBJ) "ellos")
+ ((F-IS 'OBJ) (G-FEAT 'CLITIC) "los")
+ ((F-IS 'INDOBJ) (G-FEAT 'CLITIC)
+ ((F-IS 'PP) "ellos")
+ ((F-IS 'OBJ) (M-SEG 'PRON)) "se")
-> ("les"))
-> ("ellos"))
```

anybody PRON

```
+ ((P-PSMODS '(RELCL INFCL PTPRTCL PRPRTCL PP)) "cualquiera")
+ ((P-PSMODS 'PREDADJ) "cualquier")
+ ((C-ASPECT 'NEG) "nadie")
-> ("alguien"))
```

Se puede ver fácilmente que las condiciones son de naturaleza completamente distinta y, mientras la determinación de la forma correcta en el caso de los personales podría pensarse en hacerla en generación, el tipo de operaciones que hay que realizar sobre el resto de los pronombres parece estar más relacionada con la transferencia.

En el caso del ejemplo, al evaluarse la entrada de *"they"*, la mayoría de las condiciones que se contemplan tienen que ver con la función que asume el pronombre en la oración. Si está desempeñando la función de sujeto, entonces la traducción es el pronombre personal sujeto en español; si desempeña la de objeto directo, la traducción es *"los"*, indicándose también que esta forma pronominal puede aparecer unida al verbo de manera clítica (por ejemplo: *"dá-se-los"*). Es esta última la traducción que se obtiene.

Después se accede al núcleo de N14, *"happy"*, que también tiene una única traducción, como *"feliz"*.

Y ahora sólo queda ya el PP. En éste lo primero que se traduce es, como siempre, el núcleo: *"years"*. Esta palabra no tiene ninguna característica especial, traduciéndose directamente por *"años"*. Para la preposición *"for"* tenemos sin embargo varias posibilidades, aunque desde luego no todas todavía:

for PREP

```
+ ((& (F-NOFPP 'NOUN) (M-FEAT 'TIME)) "durante")
+> ("para")
```

La primera de las condiciones es una condición compuesta, es decir, el resultado será afirmativo si y sólo si son afirmativas simultáneamente todas las decisiones agrupadas por el '&', que en este caso son dos: (1) que el núcleo de la oración preposicional sea un nombre y (2) que dicho nombre tenga la característica TIME (TiME), marca de nombres que denotan unidades o etapas de tiempo. Mediante la primitiva F-NOFPP situamos el TESTNODE en el núcleo del PP *"years"* (aunque no definitivamente, dado el tipo de control), y mediante M-FEAT comprobamos que éste tiene la marca requerida (ver N18 y N20 en la Figura 41 en la página 149). Al ser verdad ambas condiciones *"for"* recibe inmediatamente la traducción *"durante"*.

El único nodo que nos quedaría por visitar es el correspondiente al punto final de la oración, pero los signos de puntuación no se consideran de manera específica por el momento, de modo que con el acceso al diccionario ya ha finalizado lo que es la fase de transferencia léxica propiamente dicha. Por las características de esta frase no es necesario realizar ningún tipo de transformación puramente estructural independiente del léxico, transformaciones que se llevarían a cabo mediante el proceso de transferencia estructural que mencionamos anteriormente. Por lo tanto, comienza ya la fase de generación.

Es principalmente para esta fase para la que se seleccionaron algunos de los rasgos de los registros que aparecían en la Figura 40 en la página 148 y en la Figura 41 en la página 149, correspondientes al árbol decorado. Éstos son el número y la persona que, en general, se mantienen en inglés y español. El género no ha sido seleccionado, ya que no aportaría ninguna información al no corresponder la taxonomía de géneros en inglés con la española. Como veremos a continuación, el género de las construcciones se resuelve con la información conseguida del diccionario de español, al acceder a él con los nombres, parte de la oración que determina el género de sus modificadores (adjetivos, artículos, etc.).

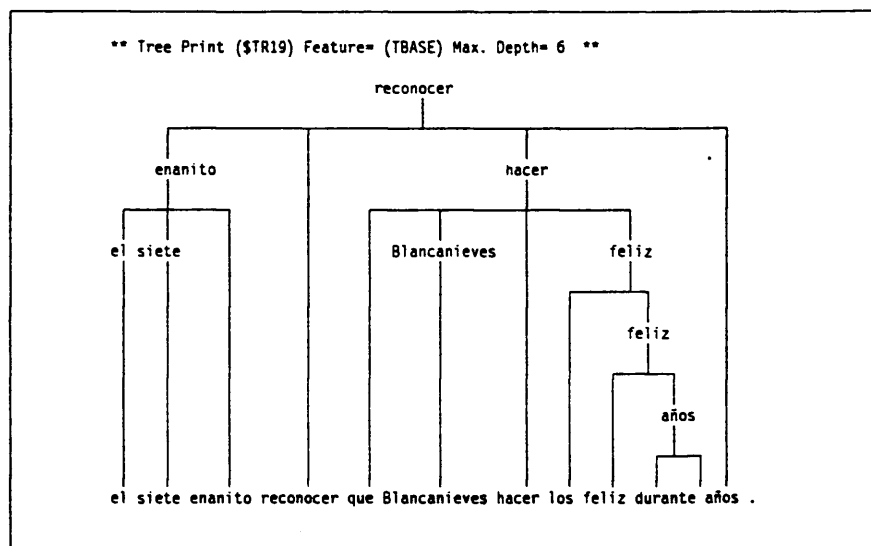


Figura 44. Transferencia Léxica: resultado final (2). "The seven dwarfs allow that Snow-White made them happy for years."

En la fase de generación tiene lugar la flexión morfológica de las palabras que aparecen como base en los distintos nodos. Para ello, como ya mencionamos en el Capítulo II de la memoria, se utiliza un diccionario del español, LEXIS, que cubre una gran parte del léxico. Recordemos que la comunicación con dicho diccionario, que se mantiene como un módulo externo, al igual que en el caso de la transferencia o el análisis, se establece por medio de una función LISP, MORPHOLOGY, a través de un procedimiento específico, MORPHOLOGY-DRIVER. El programa externo con el que se relaciona recibe una serie de parámetros que dependen de la parte de la oración de la palabra que se desea flexionar. En el caso del verbo, para hallar la forma correcta es necesario proporcionarle no sólo la base, sino también la persona, el número, el tiempo y el modo. Luego hay otros parámetros que se manejan también desde este módulo y gestionan algunas características especiales.

como son la negatividad o la presencia de clíticos. Por ejemplo, el verbo "escapar" proveniente de una oración tal como "I don't make away" se flexionará como negativo y clítico, obteniéndose después "no" "me" "escapo". De esta forma se puede manejar un único nodo durante el proceso, que contiene toda la información necesaria.

En el ejemplo, se accede al diccionario español y se flexionan correctamente las palabras. El árbol sigue teniendo la misma forma que en la Figura 44 en la página 156, pero ahora se ha incluido otro campo en los nodos traducidos, que llamamos TSTR (de 'Translated STRing'). La concatenación de los TSTR de los nodos terminales queda:

Los siete enanitos reconocen que Blancanieves hizo los felices durante años

Para el proceso de flexión morfológica ha sido necesario determinar cuáles son las relaciones de regencia entre los distintos elementos de la oración, ya que ha de existir concordancia en género, número y persona entre ellos. Estas relaciones pueden ser muy complicadas de detectar en algunas situaciones, sobre todo si existen pronombres relativos. En la Figura 45 se puede ver cuáles son en este caso (3P significa tercera persona del plural y 3S tercera persona del singular).

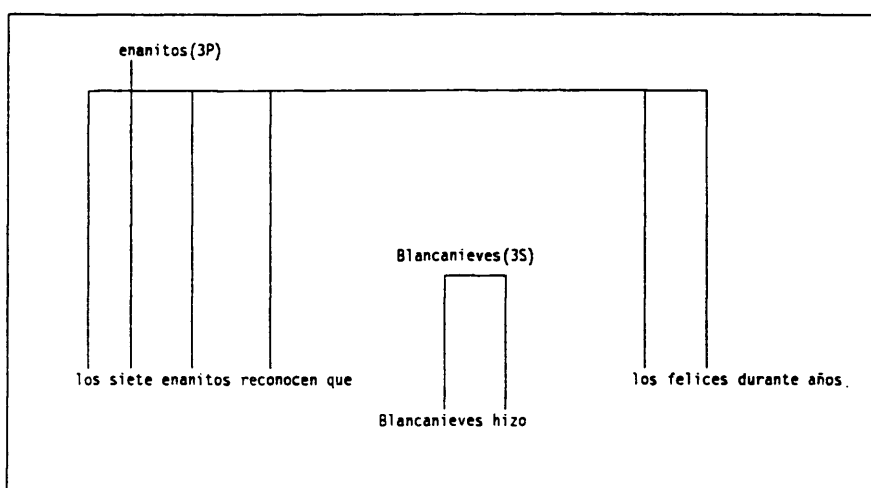


Figura 45. Regencia en la oración. "The seven dwarfs allow that Snow-White made them happy for years."

Otra de las funciones de la fase de generación es reordenar el árbol y crear o eliminar nodos para adecuar la estructura a las características del lenguaje destino, del español. En este caso lo que hay que hacer es reordenar la estructura para acomodar la situación del pronombre objeto a la

posición que tiene habitualmente en español, antes del verbo. La función LISP que lleva a cabo esta tarea se llama GENTRAN-RULE, y es capaz de reconocer ciertos patrones sobre la estructura y reordenar sus nodos adecuadamente. Existen otras dos funciones semejantes que se ocupan de los casos ditransitivos (cuando existe objeto directo e indirecto). Al finalizar esta fase, se recuperan del árbol los TSTR de los nodos terminales para formar la oración de salida. Todavía es posible que haya que realizar alguna operación más, como la compactación de los grupos "a" + "el" = "al" o "de" + "el" = "del", que son característicos de la estructura más superficial de la oración y que no deben contraerse sobre el propio árbol.

Y el resultado es:

Los siete enanitos reconocen que Blancanieves los hizo felices durante años

Para terminar este ejemplo incluimos los tiempos de ejecución (CPU) invertidos en los procesos de transferencia y generación, en la traducción de la oración propuesta. No se han realizado todavía estadísticas sobre una muestra suficientemente significativa, puesto que no se pretendía optimizar el código en esta fase del prototipo. La continuación del proyecto supondrá, por supuesto, una revisión tanto del código implementado como de los propios lenguajes de programación, ya que LISP, aunque presenta muchas ventajas para el desarrollo de prototipos, puede no ser el más adecuado para implementar sistemas de mayor envergadura.

Se ofrecen los resultados tanto para la ejecución interpretada como para la compilada, observándose en este segundo caso una mejora notable. El tiempo total de traducción (análisis + transferencia + generación), con LISP compilado y utilizando tablas Hash, es de 15.374 ms.

Figura 46. Tiempos de ejecución en ms: LISP interpretado		
	Sin tablas Hash (ms)	Con tablas Hash (ms)
Transferencia Léxica	2640	880
Transferencia	8974	4374
Generación	7998	6130
Total T + G	16972	10504

Figura 47. Tiempos de ejecución en ms: LISP Compilado		
	Sin tablas Hash (ms)	Con tablas Hash (ms)
Transferencia Léxica	1665	94
Transferencia	3873	580
Generación	6889	2794
Total T + G	10762	3374

4.3 Análisis de algunos casos con el verbo MAKE

Mediante la aplicación en MENTOR del modelo de representación de información bilingüe expuesto se ha podido resolver un conjunto de problemas de transferencia muy amplio. Aunque ya se ha comentado en algunas ocasiones cómo se pueden utilizar determinadas primitivas, pensamos que resulta interesante examinar algunos casos en detalle.

En este apartado presentamos la batería completa de oraciones que se han utilizado en las pruebas para la traducción al español del verbo "make", señalando, en los casos en que no se haya llegado a la traducción correcta, cuál es la razón. De entre estas oraciones se han seleccionado algunas que ejemplifican casos complejos o interesantes y que se analizan en mayor profundidad.

- Ejemplo 1:

(4-1) *"I allow that I made a mistake."*

(4-1t) "Yo reconozco que yo cometí un error."

- Ejemplo 2:

(4-2) *"We will make a new function."*

(4-2t) "Nosotros haremos una nueva función."

- Ejemplo 3:

(4-3) *"This table is made of wood."*

(4-3t) "Esta tabla está hecha de madera."

- Ejemplo 4:

(4-4) *"He makes \$200 a week."*

(4-4t) "Él gana \$200 por semana."

- Ejemplo 5:

(4-5) *"He made a terrible mistake."*

(4-5t) "Él cometió un error terrible."

- Ejemplo 6:

(4-6) *"You should make the previous function a new tool."*

(4-6t) "Usted debería hacer de la función previa una nueva herramienta."

- Ejemplo 7: Análisis incorrecto (ver comentarios)

(4-7) *"They will make her president."*

(4-7t) "Ellos harán su presidente."

• Ejemplo 8:

- (4-8) *"They will make him president."*
(4-8t) "Ellos le nombrarán presidente."

• Ejemplo 9:

- (4-9) *"They always make the program ready."*
(4-9t) "Ellos siempre preparan el programa."

• Ejemplo 10:

- (4-10) *"She made him happy."*
(4-10t) "Ella le hizo feliz."

• Ejemplo 11:

- (4-11) *"The circumstances made him write a new book."*
(4-11t) "Las circunstancias le obligaron a escribir un nuevo libro."

• Ejemplo 12:

- (4-12) *"This collection of functions has been made into a new program."*
(4-12t) "Este conjunto de funciones se ha convertido en un nuevo programa."

• Ejemplo 13:

- (4-13) *"If the system makes out what should be done, you may call it intelligent."*
(4-13t) "Si el sistema distingue lo que debería hacerse, usted puede llamarlo inteligente."

• Ejemplo 14:

- (4-14) *"If the system makes out the results, you may call it intelligent."*
(4-14t) "Si el sistema distingue los resultados, usted puede llamarlo inteligente."

• Ejemplo 15:

- (4-15) *"These are the components which make up the system."*
(4-15t) "Éstos son los elementos que componen el sistema."

• Ejemplo 16:

- (4-16) *"The group was made up of four people."*
(4-16t) "El grupo estuvo compuesto por cuatro personas."

• Ejemplo 17:

- (4-17) *"The car is only made up of tubes and wheels."*
(4-17t) "El coche solamente está compuesto por tubos y ruedas."

- **Ejemplo 18:**

- (4-18) *"You can't make away easily."*
- (4-18t) "Usted no puede escaparse fácilmente."

- **Ejemplo 19:**

- (4-19) *"He made for California."*
- (4-19t) "Él se dirigió a CALIFORNIA."

- **Ejemplo 20:**

- (4-20) *"If the demo makes out, the rest will be easy."*
- (4-20t) "Si la demo sale bien, el resto será fácil."

- **Ejemplo 21:**

- (4-21) *"This will make up for the previous opportunities."*
- (4-21t) "Esto compensará las oportunidades previas."

Como puede verse los resultados son altamente satisfactorios, y, siempre que el análisis sea correcto se puede asegurar una buena traducción. Por otra parte, el procedimiento interno de las primitivas se ha diseñado de modo que sea posible "recuperar" algunos errores, imprecisiones o ambigüedades del análisis mediante el reconocimiento de ciertos patrones específicos en cada caso. Esta posibilidad se comenta en detalle en el desarrollo del ejemplo 11: *"The circumstances made him write a new book"*.

Ejemplo 3: "This table is made of wood"

En este caso ha de reconocerse el uso transitivo del verbo y, por lo tanto, la existencia de un objeto directo en la estructura funcional. Sin embargo, no se verifica ninguna de las condiciones sobre el objeto o sus modificadores que se especifican en la entrada, así que nos encontramos en la traducción "por defecto" para el caso transitivo simple, que es lo correcto.

- (4-3) This table is made of wood.
- (4-3t) Esta tabla está hecha de madera.

Se ha seleccionado esta oración porque en ella se utilizan dos de las primitivas de acción del grupo de generación: G-FEAT y G-STATE, cuyo uso se pone de relieve en el ejemplo.

make VERB: "hacer"

```
+ ((F-OBJ)
-> ((G-FEAT 'NONACTIVE) (G-STATE) "hacer")))
```

Como ya se ha mencionado, la generación es la fase del prototipo que está menos estudiada y aún no se han establecido criterios detallados en todos los casos. En el apartado "2.2.3 La Generación" en la página 50 se han comentado cuáles son los criterios que se utilizaban para el tratamiento de las oraciones pasivas en el momento de la generación, siendo éstos los siguientes:

- en principio se genera "ser", salvo que se indique lo contrario,
- todas las oraciones pasivas se generan como pasivas reflejas, salvo que se indique lo contrario.

Precisamente la función que tiene G-STATE es marcar las excepciones al caso primero, y la marca NONACTIVE, incluida mediante G-FEAT, las excepciones al segundo.

Desde que se tomó esta decisión se ha experimentado, creando nuevas entradas y suele suceder que ambas funciones van generalmente asociadas. Podríamos entonces pensar en aunar sus efectos, de manera que, por ejemplo, siempre que se indique G-STATE no tenga lugar la reflexivización.

Ejemplo 6: "You should make the previous function a new tool"

Esta oración se ha elegido porque en ella se da una de las traducciones del verbo "make" que provoca el cambio de la función sintáctica de uno de los constituyentes en el proceso de transferencia léxica. Para ello se utiliza la primitiva de acción S-FCHANGE. Esta primitiva se mueve en el campo de las estructuras funcionales y no actúa directamente sobre las variables TESTNODE o WORDNODE, sino sobre los constituyentes que varían, haciendo que estos adopten la nueva función que les corresponde.

En este caso se indica que si es ésta la traducción seleccionada, el objeto directo se convierte en un complemento preposicional introducido por la preposición "de".

make VERB: "hacer"

```
((F-OBJ)
((P-NP)
-> ((S-FCHANGE '((OBJ PREPC "de"))) "hacer")))
```

En los árboles de análisis (siguiente figura) podemos ver que PEG nos proporciona dos árboles de análisis posibles, siendo el correcto el segundo de ellos. La razón es que en el primero se ha considerado "function" como el infinitivo del verbo, mientras que en el segundo se trata como nombre. En el original inglés la segunda lectura es claramente la preferida, aunque la primera pueda ser sintácticamente correcta. Por lo tanto nos encontramos con un caso en el que es necesario introducir más conocimientos lingüísticos en la fase de análisis para poder llegar a determinar el sentido correcto de la oración. Las traducciones obtenidas en ambos casos, con los dos análisis, son :

- (4-6) You should make the previous function a new tool.
 (4-6t1) Usted debería obligar al anterior a funcionar una nueva herramienta.
 (4-6t2) Usted debería hacer de la función previa una nueva herramienta.

You should make the previous function a new tool .									
DECL1	NP1	PRON1*	"you"						
	VP1	VERB1*	"should"						
	VERB2*	"make"							
	NP2	DET1	ADJ1*	"the"					
		ADJ2*	"previous"						
	?	INFCL1	VERB3*	"function"					
		NP3	DET2	ADJ3*	"a"				
			AJP1	ADJ4*	"new"				
			NOUN1*	"tool"					
	PUNC1	"."							
DECL2	NP4	PRON1*	"you"						
	VP1	VERB1*	"should"						
	VERB2*	"make"							
	NP5	DET3	ADJ1*	"the"					
		AJP2	ADJ2*	"previous"					
		NOUN2*	"function"						
	NP3	DET2	ADJ3*	"a"					
		AJP1	ADJ4*	"new"					
		NOUN1*	"tool"						
	PUNC1	"."							

Figura 48. Ejemplo 8: árboles de análisis proporcionados por PEG. "You should make the previous function a new tool."

Ejemplos 7/8: "They will make her/him president"

Estas oraciones se han escogido para poner de manifiesto algunos casos en los que se puede llegar a traducciones erróneas, no por causa del proceso de transferencia léxica, ni de la definición de una entrada, sino por problemas en el análisis que a primera vista no son evidentes. Esa es la situación que se nos presentó, después de definir y comprobar la entrada de "make", cuando en la batería de prueba se nos ocurrió utilizar el pronombre femenino "her" en vez de nombres propios o pronombres masculinos como se había hecho manualmente.

El resultado nos sorprendió porque no existía ninguna razón para esa diferencia, pero la respuesta es relativamente sencilla: este caso es ambiguo ya que "her" puede ser tanto adjetivo posesivo como pronombre personal, y ambas posibilidades dan lugar a oraciones con sentido. Por otra parte, PEG da como más probable el análisis en el que "her" actúa como adjetivo, así que de la oración 'en femenino' obteníamos:

- (4-7) They will make her president.
- (4-7t) Ellos harán su presidente.

que, aunque queda extraño, resulta sintácticamente correcto.

De la oración "en masculino" obteníamos el resultado correcto:

- (4-8) They will make him president.
- (4-8t) Ellos le nombrarán presidente.

Los análisis obtenidos pueden verse en la Figura 49 en la página 167.

La oración (4-7) es en cierto modo ambigua, aunque está clara cuál es la lectura preferida o más probable. Para poder resolver esta ambigüedad serían necesarias una de dos:

1. que PEG diera como preferido el segundo análisis,
2. que se permitiera transferir no sólo el primero, sino todos los árboles disponibles, seleccionando luego el preferido, que debería ser asimismo el segundo.

En cualquier caso, lo lógico es que se de la primera solución; esto implica que el diccionario monolingüe utilizado durante el análisis ha de enriquecerse no ya sólo con rasgos semánticos sino con algún tipo de medida "de peso" que se adjudicase a determinadas ocurrencias o combinaciones de palabras.

They will make her president .					
DECL1	NP1	PRON1*	"they"		
	VP1	VERB1*	"will"		
	VERB2*	"make"			
	NP2	DET1	ADJ1*	"her"	
		NOUN1*	"president"		
	PUNC1	"."			
DECL2	NP3	PRON1*	"they"		
	VP1	VERB1*	"will"		
	VERB2*	"make"			
	NP4	PRON2*	"her"		
	NP5	NOUN1*	"president"		
	PUNC1	"."			
They will make him president .					
DECL1	NP1	PRON1*	"they"		
	VP1	VERB1*	"will"		
	VERB2*	"make"			
	NP2	PRON2*	"him"		
	NP3	NOUN1*	"president"		
	PUNC1	"."			

Figura 49. Ejemplos 7/8: árboles de análisis proporcionados por PEG. "They will make her/him president."

Ejemplo 11: "The circumstances made him write a new book"

Esta oración se ha elegido para ejemplificar cómo mediante la adecuada elección y definición de las primitivas del diccionario bilingüe se pueden resolver algunos problemas que tienen que ver con el análisis del inglés. En esta ocasión se trata de un problema de ambigüedad que PEG aún no sabe cómo resolver, como es el nivel al que se relacionan las oraciones de infinitivo, INFCL, con la oración principal. Esta ambigüedad se puede observar en la Figura 50 en la página 168.

En esta oración es importante el uso transitivo del verbo y también la presencia de una INFCL postmodificándolo, de manera que, definiendo así la entrada, el lexicógrafo no necesita ser consciente del problema de análisis que hemos referido, encargándose P-PSMODS de considerar las distintas posibilidades. Es decir, el conocimiento/dependencia de la gramática inglesa se introduce en el código de las primitivas, no en lo que "ve" el usuario.

the circumstances made him write a new book .									
DECL1	NP1	DET1	ADJ1*	"the"					
		NOUN1*	"circumstances"						
	VERB1*	"made"							
	NP2	PRON1*	"him"						
	?	INFCL1	VERB2*	"write"					
			NP3	DET2	ADJ2*	"a"			
				AJP1	ADJ3*	"new"			
				NOUN2*	"book"				
	PUNC1	"."							

Figura 50. MAKE: análisis ambiguo. "The circumstances made him write a new book"

make VERB: "obligar a"

```
((F-0BJ)
+ ((P-PSMODS 'INFCL)(S-INFNP) (G-FLEX '(VERB 0)) '("obligar" "a"))
```

La traducción que se obtiene es la correcta:

- (4-11) The circumstances made him write a new book.
 (4-11t) Las circunstancias le obligaron a escribir un nuevo libro.

Ejemplo 12: "This collection of functions has been made into a new program"

Con este ejemplo lo que se pone de manifiesto es cómo los complementos circunstanciales, de características muy diferentes a los complementos que se han examinado hasta ahora (objetos directos, p.e.), pueden también determinar las traducciones de otros elementos de la oración. En este caso, el factor discriminante es la presencia de un complemento preposicional, detectado mediante la primitiva F-CC.

make VERB: "convertir"

```
((F-0BJ)
+ ((F-CC 'PP '(into "en"))(G-FEAT 'CLITIC) "convertir")
```

Así podemos traducir correctamente esta ocurrencia de "make" como "convertir", en vez de asignar la traducción por defecto que es "hacer":

- (4-12) This collection of functions has been made into a new program.
(4-12t) Este conjunto de funciones se ha convertido en un nuevo programa.

Ejemplo 17: "The car is only made up of tubes and wheels."

Esta oración se ha elegido para explicar cómo se abordan mediante el formalismo expuesto los casos de verbos frasales transitivos. Este tipo de verbos es de alguna manera similar a los preposicionales transitivos (en los que la preposición introduce el objeto directo) y, aunque en algún momento se pensó en utilizar primitivas diferentes para ambos, llegamos a la conclusión de que esto no era necesario, puesto que implicaría que el lexicógrafo que introduzca las entradas del diccionario siempre fuese consciente de las características (frasales o preposicionales) del verbo, originando a veces complicaciones innecesarias. Pensamos entonces que podría bastar con proporcionar las primitivas adecuadas para que se pudiese determinar la actuación del verbo en los casos transitivos, cuando además el verbo viene acompañado de una preposición o partícula en general, frente a los transitivos puros.

Se dispone para esto de la primitiva F-OBJ. F-OBJ puede admitir un parámetro que indique qué partícula o preposición se requiere que acompañe al objeto directo.

make VERB: "componer"

```
+ ((F-OBJ '(up **))  
  ((C-ASPECT '(PASSIVE))  
   (G-FEAT 'NONACTIVE)(G-STATE)(G-CC 'PP '(of "per")) "componer"))
```

En el caso de que se proporcione el parámetro en la llamada, F-OBJ internamente considera las dos posibilidades: frasal o preposicional, en función de las características de los árboles o grafos que crea el analizador. En la versión utilizada de PEG, el análisis para los frasales crea un nodo ligado mediante los registros al verbo, pero con entidad propia. Para los preposicionales, lo que se crea es un PP en el que la preposición es la que está relacionada con el verbo, y dicha relación se refleja también a través de los registros. Sin embargo, con otro analizador podrían darse otras posibilidades. Por ejemplo, podría no aparecer explícitamente ninguno de los nodos (preposición o partícula) y sólo mantenerse algún indicativo en el registro del verbo, que indicara esta situación. En la Figura 51 en la página 170 se indica, en el registro de la oración, dónde aparece explícitamente la indicación de partícula.

the car is only made up of tubes and wheels .				
<hr/>				
DECL1	NP1	DET1	ADJ1*	"the"
		NOUN1*	"car"	
	VP1	VERB1*	"is"	
	AVP1	ADV1*	"only"	
	VERB2*	"made"		
	PP1	PTC1*	"up"	
	PP2	PREP1	"of"	
		NP2	NOUN2*	"tubes"
		CONJ1*	"and"	
		NP3	NOUN3*	"wheels"
	PUNC1	","		
<hr/>				
SEGTYPE	'SENT'			
SEGTYPE2	'DECL'			
STR	" the car is only made up of tubes and wheels"			
COPYOF	VP2	"the car is only made up of tubes and wheels"		
BASE	'MAKE'			
...				
INDIC	SING PRES PASSIVE CAUSTO CMPXTRAN PERS3			
SENTYPE	'OECL'			
TOPIC	NP1 "the car"			
----> PARTICL	PREP2 "up"			
NODENAME	'DECL1'			

Figura 51. MAKE frasal: indicación en los registros. "The car is only made up of tubes and wheels"

En este caso también se especifica cómo ha de traducirse la preposición "of" si aparece introduciendo algún postmodificador: en vez de traducirse por la preposición española "de" ha de traducirse por "por". Aunque la traducción de la preposición podría también pensarse en hacerla desde la propia preposición, parece más lógico realizarla cuando nos encontramos en el verbo, ya que en este caso la traducción de dicha preposición viene determinada por la traducción concreta del verbo. El algoritmo que controla el proceso de acceso al diccionario no intentará traducir posteriormente aquellos constituyentes que ya hayan sido tratados.

La traducción que se obtiene es:

(4-17) The car is only made up of tubes and wheels.

(4-17t) El coche solamente está compuesto por tubos y ruedas.

Ejemplo 18: "You can't make away easily"

Este caso es parecido al anterior en cuanto que el verbo tiene igualmente características frasales, sólo que ahora no actúa como transitivo.

Frente a los casos transitivos que, como ya vimos, están centralizados tanto para preposicionales como para frases mediante la primitiva F-OBJ, tenemos el caso de los frases intransitivos, que se definen a través de la primitiva P-PTC, que requiere una partícula como parámetro.

make VERB: "escapar"

+ ((P-PTC '(away **)) (G-FEAT 'CLITIC) "escapar")

Mediante la introducción de la característica CLITIC se indica que esa forma del verbo es reflexiva, y como tal ha de generarse en la última fase del proceso de traducción.

Si PEG dispone de la información correspondiente que indique que esa partícula puede ir asociada al verbo, entonces en los registros encontramos la misma información que en el caso anterior (indicación de la partícula) y el árbol de análisis es también similar. Pero también puede suceder que PEG no disponga de la información necesaria y, por lo tanto, no pueda identificar la partícula (en esta oración el adverbio "away") como tal. Para poder abordar estas situaciones se han introducido criterios heurísticos, que permiten que la primitiva P-PTC no sólo detecte la estructura estándar (PP (PTC)) que se veía en el ejemplo anterior, sino otro tipo de constituyentes que son también válidos desde el punto de vista funcional. Este es el caso que se nos presenta en esta oración, donde "away" se analiza como (AVP (ADV)).

You can't make away easily .

DECL1	NP1	PRON1*	"you"
	VP1	VERB1*	"can't"
	VERB2*	"make"	
	AJP1	ADJ1*	"away"
	AVP1	ADV1*	"easily"
	PUNC1	". "	

Figura 52. MAKE frasal: aproximación. "You can't make away easily"

- 172 -

P-PTC identifica la partícula y la traducción que se obtiene es la correcta, siendo el mecanismo interno transparente al lexicógrafo:

- (4-18) You can't make away easily.
- (4-18t) Usted no puede escaparse fácilmente.

4.4 Entradas y Primitivas: Evolución

El total aproximado de entradas que hay en estos momentos en el diccionario bilingüe es de 1.200. En la Figura 53 en la página 174 se resumen algunos resultados clasificados por parte de la oración: número de entradas, porcentaje sobre el total, subconjunto de éstas que son entradas complejas y su porcentaje dentro de cada grupo (entendemos por "complejas" aquellas entradas en las que es necesario utilizar alguna primitiva para obtener traducciones correctas).

Aunque el diccionario se puede considerar, pequeño todavía, los resultados que se presentan nos dan idea de cuál será probablemente su evolución. En principio, es de suponer que la mayoría de las entradas que contenga sean nombres. Éstos serán simples en su mayoría, y no porque no existan nombres ambíguos (homógrafos, p.e.), sino más bien porque los que existen son en gran medida irresolubles en el momento actual, ya que dependen fuertemente del contexto semántico global del texto y no sólo del entorno de la oración. En muchos casos, desambiguar entre uno u otro significado depende del conocimiento del mundo de que esté provisto el sistema.

Los nombres que sí se pueden desambiguar actualmente son aquellos en los que la clave de los distintos sentidos se puede encontrar en el entorno sintáctico: pertenencia o no a PPs de un tipo u otro, características de sus premodificadores y/o postmodificadores, etc. De esta manera se ha podido abordar en muchos casos la traducción de palabras tan complicadas como "time" o "way".

La situación de los adjetivos es similar a la de los nombres, dado que muchas veces la traducción de éstos depende de la de los propios nombres a los que modifican, y hay que conseguir que ambas sean coherentes. Las primitivas que se utilizan en las entradas de adjetivos son principalmente de tipo G (Generación), y sirven para indicar características tales como el orden preferido (antes o después del nombre) o tipo de flexión, en el caso de que se transfieran como un complemento del nombre (p.e. "return parameters" que se traduce como "parámetros de retorno"). Cuando la traducción depende del nombre al que acompañan se utiliza la primitiva F-NOFADJ y, una vez en el nombre, puede hacerse uso de todas las habituales para esta parte de la oración.

Tenemos por otra parte el conjunto de adverbios, que puede considerarse prácticamente cerrado. En los adverbios complejos las primitivas que se utilizan son acciones de tipo S (Estructurales), S-TBASES concretamente, y algunas de tipo G (Generación). En la mayoría de los casos se trata de indicar las dos posibilidades (terminación en "-mente" o perífrasis adverbial) y el orden preferido a la hora de la generación (antes o después del verbo o del adjetivo). Las transferencias de conjunciones no ha sido estudiada exhaustivamente, pero es de esperar que no sea muy problemática; nuestra experiencia indica que el conjunto que manejamos actualmente es ya bastante amplio, y las traducciones correctas.

Las preposiciones y los verbos son quizá las partes de la oración más complejas. En verbos se ha trabajado intensamente y, hoy por hoy, el conjunto de primitivas desarrollado resuelve ade-

cuadramente la mayoría de los casos, aunque, por supuesto, será necesario desarrollar otras que resuelvan casos que aún no se nos hayan presentado. Las preposiciones son también enormemente complejas, habiéndose trabajado menos con ellas que con los verbos; aunque se han desarrollado ya dos primitivas específicas que deciden sobre el núcleo del PP donde se incluye la preposición, F-NOFPP, y el constituyente al que afecta dicho PP, F-MODHEAD. Como muchas de las traducciones de las preposiciones tienen que ver con alguno de estos dos elementos, existen ya casos complejos que se pueden abordar con cierta facilidad (ver las descripciones de "in" y "to" en el Apéndice D, "Algunas entradas completas" en la página 209).

Figura 53. Resultados sobre las entradas						
	V	(Pro)N	Adj	Adv	C	Prep
Entradas	176	438	291	151	36	46
% Total	15%	38%	26%	13%	3%	4%
Complejas	39	25	34	80	1	24
% Complejas PO	22%	6%	12%	53%	3%	52%

En cuanto al tiempo que se utiliza en la definición de una nueva entrada, éste varía considerablemente dependiendo de su complejidad, y puede oscilar entre unos segundos, para las traducciones inmediatas o uno-a-uno, y hasta 10 o 12 horas para las más complejas de las estudiadas. Dentro de las complejas existen tipos, es decir, hay entradas que internamente tienen la misma estructura; lo que significa que se puede utilizar el "esqueleto" ya pre-hecho para únicamente rellenarlo con la información de la nueva voz. De esta manera es de prever que se gane tiempo en las actualizaciones. Un ejemplo típico de esta situación son los de verbos que sólo tienen un uso transitivo simple y uno intransitivo, y que en español dan lugar a su vez a un transitivo y a un intransitivo pronominal (p.e. "devolver" y "volver(se)" de "return").

Para finalizar el conjunto de resultados hemos resumido en la siguiente figura cuáles son las primitivas que se han utilizado de hecho en las entradas, según la parte de la oración. Con esta tabla se proporciona una información de naturaleza empírica, que es útil para la realización de nuevas actualizaciones o para profundizar en cuáles son los factores de discriminación de los distintos sentidos de las palabras.

Figura 54. Primitivas Utilizadas

Verbos	Preposiciones	(Pro)Nombres	Adjetivos	Adverbios
C-ASPECT F-CC F-INDOBJ F-OBJ F-SUBJ P-AJP P-NP P-PRPRCL P-PSMODS P-PTC P-TYPE M-FEAT M-IS	C-ASPECT F-MODHEAD F-NOFPP F-VERB P-PRMODS M-FEAT M-IS M-MORPH	C-ASPECT C-TYPE F-IS F-OBJ F-VERB P-PRMODS P-PSMODS M-IS M-SEG	F-NOFADJ M-FEAT M-IS	C-ASPECT
S-COMPL S-FCHANGE S-INFNP S-XCHANGE G-CC G-FEAT G-FLEX G-SETPREP G-STATE	S-INFNP G-FLEX	S-INFNP G-SETVERB G-FEAT G-FLEX	G-FEAT G-FLEX S-RELCL	G-FEAT G-FLEX S-TBASES

Conclusiones

El trabajo que se ha descrito en la presente memoria se enmarca dentro de un proyecto de construcción de un prototipo multilingüe de traducción automática que comenzó en 1987 y que se proponía abordar la traducción de inglés a cuatro lenguas meta: español, hebreo, portugués y finés. El dominio de las traducciones estaba en principio restringido a textos técnicos de informática (manuales de IBM), y el trabajo particular sobre cada una de las lenguas meta se realizaría en cada país independientemente, siendo uno de los objetivos prioritarios la experimentación exhaustiva en la representación y determinación de las equivalencias léxicas entre los distintos idiomas.

Este prototipo implementa el modelo de transferencia, como muchos de los sistemas que se encuentran actualmente en estado de comercialización o desarrollo (METAL de Siemens, EUROTRA de la CEE, etc...), donde la fase de análisis del idioma origen es compartida y única para todas las lenguas terminales, que hacen uso del mismo analizador y del mismo diccionario de análisis. La fase de generación es particular de cada idioma y todavía está en un estado muy experimental en todos los grupos que trabajan en el proyecto.

En cuanto a la fase de transferencia, en cada grupo se ha investigado sobre distintos modelos y mecanismos, de los cuales hemos presentado el formalismo para el diccionario bilingüe que se ha desarrollado en el Centro de Investigación UAM-IBM de Madrid para la traducción al español. En esta fase llevamos a cabo tanto la transferencia estructural como la léxica entre el par de lenguas involucrado, potenciando esta última sobre la primera mediante los mecanismos descritos, que hacen recaer gran parte de las funciones de transferencia en las definiciones de las palabras y sus equivalentes, contenidas en un diccionario bilingüe. Dicho diccionario se encuentra estructurado de acuerdo con el formalismo que se ha presentado.

Con este enfoque nos adherimos a la tendencia actual de potenciar el protagonismo de las operaciones léxicas en los modelos de traducción automática. Relegar estas operaciones ha demostrado no ser el camino más adecuado, dado que algunos de los problemas que la transferencia léxica lleva asociados están íntimamente relacionados con transformaciones estructurales profundas en la oración, transformaciones que no es conveniente realizar a posteriori, puesto que están realmente dirigidas por el léxico. En este sentido, el formalismo que se ha presentado en esta memoria proporciona un marco adecuado para el desarrollo de las definiciones de los equivalentes léxicos entre palabras de diferentes idiomas, y está basado en el trabajo conjunto de un grupo interdisciplinar en lingüística computacional, que ha examinado cuidadosamente los puntos de vista tanto de lexicógrafos como de informáticos.

Al igual que en otros formalismos, para la descripción de las entradas se utiliza un modelo basado en listas (tipo LISP). En nuestro caso se definió así por razones de comodidad, ya que éste era el entorno de programación en el que estábamos trabajando. Esta decisión no obliga o presupone que el modelo de diccionario así definido sólo pueda ser utilizado en este entorno, ya que el

formalismo *no es dependiente del lenguaje LISP*, y los datos ya definidos (las entradas del diccionario) pueden ser fácilmente utilizados en otros entornos de programación. Para ello es necesario modificar el intérprete que se ha escrito para acceder a los datos, de manera que la expresión final de la entrada resulte la requerida en cada caso. Las primitivas que se ejecutan al invocarse las claves correspondientes en la definición son absolutamente independientes de la propia definición, por lo que es posible cambiarlas en función de:

1. las características del entorno de trabajo (si se quiere implementar el modelo en otros entornos de programación),
2. las particularidades del analizador sobre cuyos resultados actúan (cambio el analizador o diferentes versiones del utilizado).

En este sentido, se resaltan las características perfectamente modulares del formalismo de definición de información bilingüe presentado, características que han contribuido sobremedida a demostrar su idoneidad para el desarrollo de sistemas complejos de PLN, como es el programa de traducción automática que se ha llevado a cabo.

Desde el punto de vista de los lexicógrafos, el sistema de claves y sintaxis que se ha definido resulta ser sencillo de utilizar, puesto que se ha hecho especial hincapié en la precisión de los conceptos que las primitivas implementan. Por otra parte, la sintaxis definida por los símbolos de control es suficientemente sencilla, y a la vez versátil, para ser utilizada por usuarios no familiarizados con la programación. Responder adecuadamente a estas dos necesidades era uno de nuestros objetivos, ya que, una vez diseñados los mecanismos de declaración, y su reflejo implementado en las funciones asociadas a las primitivas, el diccionario se convierte en una herramienta básicamente para lexicógrafos: éstos han de encargarse, como especialistas, de la creación de nuevas entradas y de la modificación de las ya existentes. Sobre todo en los momentos iniciales, el trabajo directo que enfrenta a los lexicógrafos con la herramienta diseñada es básico como fuente de detección de errores, y se constituye en el vehículo principal para conseguir una mayor adecuación de las primitivas a las necesidades, más matizadas, que se detecten. Es en esta interacción de donde han de surgir las claves nuevas que responderán a los requerimientos lingüísticos con los que todavía no nos hemos enfrentado. La determinación de dichas claves dará lugar a la definición de nuevas primitivas (de acción o decisión) que se incorporarán sin disrupciones a los conjuntos ya establecidos. De esta forma, apoyándonos en el trabajo de los lexicógrafos, la completitud de los conjuntos de primitivas para el caso de transferencia que llevamos entre manos queda garantizada.

Durante el desarrollo de esta memoria nos hemos ceñido siempre al análisis del formalismo aplicado a la transferencia léxica entre inglés y español pero, tal como está diseñado, éste resulta ser extensible a otros pares de lenguas en el sentido que se ha postulado en el Capítulo 3:

- las primitivas sobre la lengua origen (inglés) cubren gran parte de las claves necesarias para la desambiguación,
- las acciones, al estar enfocadas a la generación del español, pueden también utilizarse al transferir desde otros lenguajes.

A este respecto se ha evaluado la realización de la transferencia léxica entre el inglés y el hebreo utilizando el formalismo descrito, habiéndose comprobado que el conjunto de primitivas de decisión definido es en gran parte igualmente necesario. Esta verificación ha permitido asimismo completar dicho conjunto con nuevas claves que no son precisas para la traducción al español. Como ya se ha dicho, el que se trabaje para una LM u otra no modifica el conjunto de decisiones, pero enriquece los puntos de vista sobre la LO y, por lo tanto, contribuye a la completitud del conjunto de claves de desambiguación.

Fuera ya de las previsiones, en esta misma experiencia se examinaron las primitivas de acción descritas para el español, desde el punto de vista del hebreo. El resultado fue en cierto sentido sorprendente, ya que, aunque no se da, por supuesto, la misma problemática para una LM y otra, el tipo de problemas que nosotros habíamos detectado tienen grandes similitudes en el caso del hebreo, por lo que parte de la estructura de nuestras primitivas (salvando los detalles) es asimismo transportable. Pero la definición de este conjunto para el hebreo requiere evidentemente la inclusión de acciones que no tienen sentido en nuestro caso. Para lenguas más próximas al español, lenguas romances, es de esperar que los resultados sean aún mejores, ya que una parte de los problemas que las primitivas de acción resuelven serán directamente aplicables.

El prototipo de traducción automática en el que se enmarca nuestro trabajo, al estar todavía en periodo de investigación, presenta ciertas carencias frente a otros sistemas que están bien establecidos. En primer lugar, es necesario crear más entradas, hasta llegar a un diccionario de transferencia léxica con la envergadura suficiente como para poder ser utilizado con confianza en procesos reales de traducción. Un diccionario de estas características se calcula que ha de poseer del orden de 10.000 entradas, que es un salto de magnitud frente al tamaño del que actualmente disponemos. Para poder superar este escalón es necesario realizar un cambio cualitativo en el método de definir dichas entradas. Es decir, para que el requerimiento anterior se pueda llevar a cabo en un tiempo razonable es imprescindible profundizar en el desarrollo de las interfaces, tanto con el usuario final como con los lexicógrafos que introduzcan los datos (al estilo de LOGOS, ALPS o SYSTRAN).

También sería conveniente introducir algún tipo de codificación en las entradas, de la misma manera que los sistemas que hemos citado anteriormente. La utilización de estos códigos (área temática, frecuencia, etc.) evidentemente mejorarían el sistema, haciendo de filtro o medida de 'peso' en algunos casos.

En cuanto a la continuación de este trabajo, a partir del que se está realizando actualmente para el proyecto MENTOR, nos planteamos varias posibles perspectivas de ampliación del sistema.

En primer lugar, nos interesa estudiar la manera de realizar la extracción de la información bilingüe de manera automática. Algo se ha comentado sobre este tema a lo largo de la memoria y, aún constatando su complejidad, resulta ser una evidente tentación para este tipo de sistemas. Para ello, hemos comenzado una colaboración con el grupo de lingüística computacional del Centro de

Investigación de IBM en Yorktown Heights, fruto de la cual se dispone ya de un primer analizador automático de la Información léxica contenida en el diccionario Collins inglés-español. El siguiente paso, una vez que el analizador se complete, será adaptar la información que se extraiga mediante este mecanismo a la estructura (primitivas y símbolos de control) del formalismo que hemos presentado. Esto permitirá incrementar rápidamente el número de entradas contenidas en el diccionario bilingüe, lo que nos permitirá mejorar la calidad de los textos traducidos.

Nos planteamos también la posibilidad de incluir algún método para la gestión de los sinónimos de las traducciones de las palabras. En algunos casos, la elección de una u otra traducción es un cuestión puramente estilística que depende de las preferencias personales del traductor, por lo que la posibilidad de elegir entre varias traducciones quizá debiera dejarse a su criterio. Esta opción está claramente relacionada con el diseño de las interfaces del puesto de trabajo integrado para el traductor, que es otro de los aspectos del sistema que nos proponemos abordar.

Por otra parte, en el estado en que nos encontramos una de las primeras extensiones naturales del proyecto es la evaluación rigurosa de las posibilidades de este formalismo en cuanto a su aplicación a otros pares de lenguas. Como hemos mencionado, ya se ha evaluado con el par de lenguas inglés-hebreo con resultados excelentes. Nuestro próximo objetivo será incluir alguna de las lenguas romances, probablemente el par inglés-portugués, donde esperamos que las primitivas de decisión sean útiles en su casi totalidad y las de acción cuando menos muy semejantes.

En cuanto a las extensiones o mejoras puramente computacionales, se está considerando la posibilidad de trasvasar el sistema de traducción a estaciones de trabajo personales. Para ello es conveniente reprogramarlo, en parte o en su totalidad, en otro lenguaje de programación que sea (1) más transportable y (2) más rápido. Desde estos puntos de vista pensamos que C puede ser el candidato idóneo. Esta reprogramación no afectaría a los contenidos del diccionario, aunque supondría la nueva formulación de los programas asociados a las primitivas en el lenguaje de programación elegido y en función del analizador que se utilizase. Aún así, el trabajo no será muy costoso puesto que su número es limitado.

Para terminar, mencionaremos que se ha comenzado a experimentar con la información léxica bilingüe que hemos formalizado con vistas al diseño de un sistema de traducción con características de programación orientada a objetos. En este sistema el resultado del análisis, en vez de asimilarse a un árbol decorado, sería una estructura jerarquizada de objetos con sus propiedades características. La función '*traducir*' se va a implementar como un mensaje que, al ser recibido por el objeto, provocará el acceso al diccionario para esa palabra y la posterior evaluación de las condiciones de decisión, siendo la traducción el resultado final, que se proporcionaría también como mensaje. De todas formas, los detalles están aún por determinar, y es sólo una idea sobre la que hemos empezado a trabajar, pero que ofrece buenas perspectivas y apoya los argumentos que nos han llevado a desarrollar el formalismo para la representación y utilización de información léxica bilingüe presentado.

Referencias

- Aba89. Abarbanel R. M. y otros, *AI Tools and Techniques*, Norwood, New Jersey: Ablex Publishing Corporation, 1989.
- AlK83. Al-Kasimi A. M., "The interlingual/translation dictionary", en R.R.K. Hartmann, editor, *Lexicography: Principles and Practice*, págs. 153-162, Academic Press, Inc., 1983.
- Bal86. Balfour R. W., *Machine Translation: A Technology Assessment*, BMT Consultants, 1986.
- Bog89. Boguraev B. y Briscoe E., *Computational Lexicography for Natural Language Processing*, Essex: Longman, 1989.
- Boit86. Boitet C. y Nedobejkine N., "Towards Integrated Dictionaries for M(a)T: motivations and linguistic organisation", *Proceedings of COLING86*, págs. 423-426, Bonn, 1986.
- Boi89. Boitet C., "Motivations, Aims and Architecture of the LIDIA project", *MT Summit II*, págs. 50-54, Munich, Agosto de 1989.
- Bra79. Brachman R. J., "On the Epistemological Status of Semantic Networks", en N. V. Findler, editor, *Associative Networks: Representation and Use of Knowledge by Computers*, Academic Press, 1979.
- Byr83. Byrd R. J., "Word Formation in Natural Language Processing Systems", *Proceedings of the 8th IJCAI*, págs. 704-706, Karlsruhe, West Germany, Agosto 8-12 de 1983.
- Byr85. Byrd R. J. y Chodorow M. S., "Using an on-line dictionary to find rythming words and pronunciation for unknown words", *Proceedings of the 23rd Annual Meeting of he ACL*, págs. 277-283, 1985.
- Byr86. Byrd R. J., Neumann G., y Anderson S., *DAM - A Dictionary Access Method*, Yorktown Heights, New York: IBM Internal Report, Noviembre 1986. Draft.
- Byr87. Byrd R. J., Calzolari N., Chodorow M. S., Klavans J. L., Neff M. S., y Rizk O. A., "Tools and Methods for Computational Lexicology", *Computational Linguistics*, vol. 13, núm. 3-4, págs. 219-240, Julio-Diciembre de 1987.
- Cal88. Calzolari N. y Picchi E., "Acquisition of Semantic Information from an On-line Dictionary", *Proceedings of COLING88*, vol. 1, págs. 87-92, Budapest, Agosto de 1988.
- Cal89. Calzolari N. y Zampolli A., "Very large linguistic databases: The problem of the reusability of lexical resources", *The Dynamic Text*, págs. 47-48, Toronto: Proceedings of the 9th ICCH and the 16th ALLC, Junio de 1989.

- Car84. Lázaro Carreter F., *Diccionario de Términos Filológicos*, Madrid: Editorial Gredos S.A., 1984.
- Cas85. Casajuana R. y Rodríguez C., "Verificación ortográfica en castellano. La realización de un diccionario en ordenador", *Español Actual*, núm. 44, 1985.
- Cas87. Casajuana R., Rodríguez C., Sopeña L., y Villar C., "Towards an Integrated Environment for Spanish Document Verification and Composition", *Proceedings Third Conference of the ACL (European Chapter)*, núm. 44, Copenhagen, Abril de 1987.
- Col84. Collins, *The New Collins (English - Spanish, Español - Inglés)*, Glasgow, Scotland: Collins Publishers Ltd., 1984.
- Cul87. Cullingford R. E. y Onyshkevych B. A., "An experiment in lexicon-driven machine translation", en Sergei Nirenburg, editor, *Machine Translation: theoretical and methodological issues*, págs. 278-301, Cambridge University Press, 1987.
- EDR88. EDR, Technical Report, Japan Electronic Dictionary Research Institute Ltd., Noviembre 1988.
- Gol88. Golan I., Lappin S., y Rimón M., "An Active Bilingual Lexicon for Machine Translation", *Proceedings of COLING88*, vol. 1, págs. 205-211, Budapest, Agosto de 1988.
- Hac90. Ten Hacken P., "Reading Distinction in Machine Translation", *Working Papers in Natural Language Processing*, núm. 6, Eurotra-Leuven: Katholieke Universiteit, 1990.
- Har83. Harper K. E., "Dictionary Problems in Machine Translation", en Paul L. Gavin, editor, *Natural Language and the Computer*, McGraw Hill Book Co., 1983.
- Har83. Hartmann R. R. K., "On Theory and Practice", en R.R.K. Hartmann, editor, *Lexicography: Principles and Practice*, págs. 3-12, Academic Press, Inc., 1983.
- Hay85. Hayes P. J. y otros, *Readings in Knowledge Representation*, California: Morgan Kaufmann Publishers Inc., 1985.
- Heid72. Heidorn G. E., *Natural Language Inputs to a Simulation Programming System*, tesis doctoral, Naval Postgraduate School, Monterey, California Octubre 1972.
- Hei82. Heidorn G. E., "Experience with an easily computed metric for ranking alternative parsers", *Proceedings of the 20th Annual Meeting of the ACL*, págs. 82-84, 1982.
- Hirs89. Hirst G., *Semantic Interpretation and the Resolution of Ambiguity*, Cambridge: Cambridge University Press, 1989.

- Isa88a. Isabelle P., "Electronic Dictionaries in Canada", *Proceedings of ISED'88*, págs. 29-31, Tokyo, Noviembre de 1988.
- Isa88b. Isabelle P., "Electronic Dictionaries and MT Systems", *Proceedings of ISED'88*, págs. 32-34, Tokyo, Noviembre de 1988.
- Jen83. Jensen K. y Heidorn G., "The Fitted Parse: 100% Parsing Capability in a Syntactic Grammar of English", *Proceedings Conference on Applied Natural Language Processing*, Santa Monica, California, Febrero de 1983.
- Jen86. Jensen K., PLNLP, PEG and CRITIQUE: Three Contributions to Computing in the Humanities, Yorktown Heights, New York: IBM Research Report RC 11841, 1986.
- Jen87. Jensen K., PEG 1986: A Broad-coverage Computational Syntax of English, Yorktown Heights, New York: IBM Internal Report, Noviembre 1987.
- Kaj88. Kaji H., "Dictionary Structure for Flexible Lexical Transfer in Machine Translation", *Proceedings of ISED'88*, págs. 36-38, Tokyo, Noviembre de 1988.
- Kay86. Kay M., "Machine Translation Will Not Work", *Proceedings of the 24th Annual Meeting of the ACL*, Junio de 1986.
- Kin84. King M. y Perschke S., EUOTRA, Paper presented for the Lugano Tutorial on Machine Translation, Abril 1984.
- Klav88. Klavans J., "COMPLEX: A Computational Lexicon for Natural Language Systems", *Proceedings of COLING88*, vol. 2, págs. 815-823, Budapest, Agosto de 1988.
- Kos83. Koskenniemi K., *TWO-LEVEL MORPHOLOGY: A General Computational Model for Word-Form Recognition and Production*, Helsinki: University of Helsinki Publications, 1983.
- Lan88. Landsbergen J., "Dictionaries for Rosetta", *Proceedings of ISED'88*, págs. 45-47, Tokyo, Noviembre de 1988.
- Lar84. Larousse, *Larousse Gran Diccionario Español- Inglés, English - Spanish*, Ediciones Larousse, 1984.
- Len88. Lenders W., "Computer Aided Extraction of Canonical Lexical Information from Lexical Definitions", *Proceedings of ISED'88*, págs. 52-54, Tokyo, Noviembre de 1988.
- Lon78. Longman, *Longman Dictionary of Contemporary English*, Harlow and London, England: Longman Group Ltd., 1978.

- Mae88a. Maegaard B., "EUROTRA Dictionaries", *Proceedings of ISED'88*, págs. 58-61, Tokyo, Noviembre de 1988.
- Mae88b. Maegaard B., "Designing and testing linguistic development phases in a machine translation project", *Proceedings of COLING88*, vol. 1, págs. 387-389, Budapest, Agosto de 1988.
- Mag89. Rodriguez-Magro C., Sopeña L., y Villar C., "Caracterización sintáctica de los verbos castellanos para una gramática por ordenador", *Procesamiento del Lenguaje*, núm. 7, págs. 99-104, Enero de 1989.
- McC88. McCord M., Design of LMT: A Prolog-Based Machine Translation System, Yorktown Heights, New York: IBM Research Report RC 12455, Febrero 1988.
- Melb88. Melby A. K., "Lexical Transfer: A Missing Element in Linguistic Theories", *Proceedings of COLING86*, págs. 104-106, Bonn, 1986.
- Mel88a. Melby A. K., "Electronic Dictionaries and the Translator's Workbench", *Proceedings of ISED'88*, págs. 62-64, Tokyo, Noviembre de 1988.
- Mel88b. Melby A. K., "Lexical Transfer: Between a Source Rock and a Hard Target", *Proceedings of COLING88*, vol. 2, págs. 411-413, Budapest, Agosto de 1988.
- Mii88. Miike S. y Amano S., "AS-TRANSAC Bilingual Dictionaries", *Proceedings of ISED'88*, págs. 65-67, Tokyo, Noviembre de 1988.
- Nag88. Nagao M., "Taxonomy, Content and Construction of Electronic Dictionaries", *Proceedings of ISED'88*, págs. 68-75, Tokyo, Noviembre de 1988.
- Nak88. Nakamura J. y Nagao M., "Extraction of Semantic Information from an Ordinary English Dictionary and its Evaluation", *Proceedings of COLING88*, vol. 2, págs. 459-464, Budapest, Agosto de 1988.
- Nef88. Neff M. S. y Byrd R. J., Wordsmith User's Guide: Version 2.0, Yorktown Heights, New York: IBM Research Report RC 13411, Noviembre 1988.
- Nir89. Nirenburg S., "KBMT-89 - a knowledge based MT project at Carnegie Mellon University", *MT Submit II*, págs. 106-112, Munich, Agosto de 1989.
- Nom87. Nomiyama H., Tree Transducer, IBM Tokyo Research Laboratory, Abril 1987.
- Nyb88. Nyberg E., FRAMEKIT User's Guide. Technical Memo., Carnegie Mellon University, 1988.

- Pai88. Palmer R. F., *The English Verb (2nd edition)*, Longman, 1988.
- Pap86a.Papegaij B. C., Sadler V., y Witkam A. P. M., "Experiments with an MT-Directed Lexical Knowledge Bank", *Proceedings of COLING86*, págs. 432-434, Bonn, 1986.
- Pap86b.Papegaij B. C., Word expert semantics: an interlingual knowledge-based approach (DLT 1), Dordrecht: Foris, 1986.
- Pet80. Peterson J. E., *Word Sense Selection in a One-to-Many, Interactive Computer-Assisted Translation System*, Brigham Young University, 1980.
- Pig88. Pigott I. M., *Systran machine translation at the EC Commission: present status and history*, Luxemburgo: CCE, 1988.
- Qui85. Quirk R., Greenbaum S., Leech G., y Svartvik J., *A Comprehensive Grammar of the English Language*, Longman, 1985.
- Red88. Redondo T., Rodríguez P., Sopeña L., y Zapata I., MENTOR88/S: The Spanish Prototype (draft), Madrid: Internal Report, Junio 1988.
- Red89a.Redondo T., Rodríguez P., Sopeña L., y Zapata I., "Tratamiento de algunas estructuras complejas en un Sistema de Traducción Automática", *Procesamiento del Lenguaje*, núm. 7, págs. 181-188, Enero de 1989.
- Red89b.Redondo T., Rodríguez P., Sopeña L., y Zapata I., "Transferencia Léxica en un Prototipo de Traducción Automática", *Procesamiento del Lenguaje*, núm. 8, págs. 7-21, Diciembre de 1989.
- Riz89. Rizk O. A., Sense Disambiguation of Word Translations in Bilingual Dictionaries: Trying to Solve the Mapping Problem Automatically, Yorktown Heights, New York: IBM Research Report RC 14668, Junio 1989.
- Rod87. Rodríguez P. y Sopeña L., "MENTOR: un proyecto de traducción automática", *Procesamiento del Lenguaje*, núm. 5, págs. 99-105, Octubre de 1987.
- Rod89. Rodríguez P. y Sopeña L., "Un prototipo multilingüe de traducción asistida por ordenador", *Revista de la AEIA*, vol. XXII, núm. 1, págs. 46-51, Marzo de 1989.
- Shi89. Shigeru S., "Practical Experience in the Application of MT Systems", *MT Submit II*, págs. 125-127, Munich, Agosto de 1989.
- Sch88. Schubert K., "Linguistic and extra-linguistic knowledge", *Computers and Translation*, núm. 1 (3), págs. 125-152, 1988.

- Slo85. Slocum J., "A Survey of Machine Translation: Its History, Current Status and Future Prospects", *Computational Linguistics*, vol. 11, núm. 1, págs. 1-17, January-Marzo de 1985.
- Sop88. Sopeña L., "Traducción por ordenador: situación, problemas y perspectivas", *Novatica*, vol. XIV, núm. 73, págs. 30-37, Abril de 1988.
- Sop89. Sopeña L., "Nuevas perspectivas de la traducción", *Lingüística Aplicada*, Abril de 89.
- Sow84. Sowa J. F., *Conceptual Structures: Information Processing in Minds and Machines*, Addison-Wesley Publishing Co., 1984.
- Sto65. Stockwell R. P., Donald Bowen J. D., y Martin J. W., *The Grammatical Structures of English and Spanish*, Chicago: The University of Chicago Press, 1965.
- Tho86. Thomson A. J. y Martinet A. V., *A Practical English Grammar (4th edition)*, Oxford University Press, 1986.
- TSS87. TSS, Translation Support System, Provo, Utah: Automated Language Processing Systems Inc., 1987.
- Tsuj88. Tsujii J. y Ashizaki T., "The Dictionaries of the MU II Project", *Proceedings of ISED'88*, págs. 85-87, Tokyo, Noviembre de 1988.
- Tsut86. Tsutsumi T., "A Prototype English-Japanese Machine Translation System for Translating IBM Computer Manuals", *Proceedings of COLING86*, págs. 646-651, Agosto de 1986.
- Tuc84. Tucker A. B., "A Perspective on Machine Translation: Theory and Practice", *Communications of the ACM*, vol. 27, núm. 4, págs. 322-329, Abril de 1984.
- Uch89. Uchida H., "ATLAS", *MT Summit II*, págs. 152-157, Munich, Agosto de 1989.
- Van83. VanSlype G., Guinet J. F., Seitz F., y Benejam E., *Better Translation for Better Communication*, Oxford: Pergamon, 1983.
- Vas85. Vasconcellos M. y León M., "SPANAM and ENGSPAN: Machine Translation at the Pan American Health Organization", *Computational Linguistics*, vol. 11, núm. 2-3, págs. 122-136, Abril-Septiembre de 1985.
- Wal87. Walker D., Calzolari N., y Zampolli A., *Towards a polytheoretical lexical database*, Pisa: ICL, 1987.

- Wal90. Walker D., Calzolari N., y Zampolli A., *Automating the Lexicon: Research and Practice in a Multilingual Environment*, Cambridge: The MIT Press, 1990. Próxima edición.
- Ward88. Ward N., "Issues in Word Choice", *Proceedings of COLING88*, vol. 2, págs. 726-731, Budapest, Agosto de 1988.
- Warn82. Warneson I. y Moreau R., Recherche algorithmique de la pertinence synonymique par agrégation de similarités, IBM France, Octubre 1982.
- Web67. Webster, *Webster's Seventh New Collegiate Dictionary*, Chicago, Illinois: Rand McNally and Co., 1967.
- Whe85. Wheeler P., "LOGOS", *Sprache und Datenverarbeitung*, núm. 9, págs. 11-21, 1985.
- Whi87. White J. S., "The research environment in the METAL project", en Sergei Nirenburg, editor, *Machine Translation: theoretical and methodological issues*, págs. 225-248, Cambridge University Press, 1987.
- Wil88. Wilks Y., Fass D., Guo Ch., McDonald J. E., Plate T., y Slator B. M., "Machine Tractable Dictionaries as Tools and Resources for Natural Language Processing", *Proceedings of COLING88*, vol. 2, págs. 750-755, Budapest, Agosto de 1988.
- Woo87. McGee Wood M., Pollard E., Horsfall H., Holden N., Chandler B., y Carroll J., "Dictionary Organisation for Machine Translation: The Experiences and Implications of the UIMST Japanese Project", *3rd ACL European Chapter*, 1987.
- Zap89. Zapata I., BIDICT Access & Update Interface (draft), Madrid: Internal Report, Septiembre 1989.

Apéndices

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2

Con la información contenida en los apéndices que se presentan a continuación se pretenden cubrir algunos temas que han surgido a lo largo de esta exposición y cuya inclusión en la misma habría desviado en cierta forma la línea general del discurso.

En el primer apéndice se detalla la notación empleada en la descripción de las primitivas en el Capítulo 3.

El segundo, *Marcas Semánticas*, es parte del trabajo de Teófilo Redondo Pastor dentro de nuestro grupo, y se ha incluido por ser esta estructura de marcas semánticas sobre la que se vienen definiendo las entradas bilingües.

En el tercer apéndice se muestra un conjunto de entradas seleccionadas entre las que el sistema maneja actualmente. Se han escogido palabras con partes de la oración diferentes, algunas sencillas y otras más complejas, que dan una visión general del aspecto de las entradas expresadas según el formalismo propuesto. En particular, se incluyen prácticamente todas las entradas de las que se hace mención a lo largo de la memoria.

Y, para finalizar, en el cuarto apéndice se incluyen algunos datos de los programas que constituyen el sistema (líneas de código).

Apéndice A. Notación

Los parámetros se pueden escribir tanto en mayúsculas como en minúsculas, salvo aquellos que aparecen entre comillas, que deben aparecer obligatoriamente en minúsculas.

base-espa Cualquier forma base posible de una palabra española.

Valores: Cualquiera.

base-ingl Cualquier posible forma base de una palabra inglesa.

Valores: Cualquiera.

c-aspecto Una de las características que determinan el aspecto de una oración. Utilizado en C-ASPECT.

Valores: PRES, PAST, FUTURE, PERFECT, NOPERFECT, PROG, NOPROG, PASSIVE, NOPASSIVE, NEG y NONEG.

c-generación Característica utilizada en la fase de generación.

Valores: CLITIC, NEG, NOPRON, NONACTIVE, PRENOUN, SGENDER and SNUMBER.

c-morfológica Marca morfológica que caracteriza a un elemento de la oración.

Valores: SING, PLUR, MALE, FEMALE, NEUTER, PERS1, PERS2, PERS3, COMP, SUPR, INDEF y DEF.

comp-espa Complemento español que surge durante la transferencia.

Valores: NOUNC.

función-s Función sintáctica de un elemento de la oración.

Valores: SUBJ, OBJ, INDOBJ, VERB, PP y AVP.

lista-c-aspecto Lista de características de aspecto.

Valores: Una lista compuesta de valores posibles de c-aspecto.

lista-c-generación Lista de características de generación.

Valores: Una lista compuesta de valores posibles de c-generación.

lista-cambios-función Lista de pares de cambios funcionales entre el par de lenguajes considerado.

Valores: Una lista compuesta por uno o varios de los siguientes pares: ('OBJ 'PREPC "prep-espa"), ('OBJ 'SUBJ), ('SUBJ 'INDOBJ) y (PREPC INDOBJ).

lista-flex Lista de 'O's y/o Partes de la Oración (POS).

Valores: 'O', VERB, NOUN, ADJ.

lista-tipos-psmod Lista de posibles tipos de postmodificadores verbales.

Valores: Una lista compuesta por valores posibles de tipo-psmod.

marca-semántica Una de las marcas semánticas utilizadas en el sistema.

Valores: Ver el Apéndice B, "Marcas Semánticas" en la página 199.

modo Modo verbal.

Valores: INDICativo y SUBJuntivo.

p-espa Preposición o adverbio español.

Valores: todas las preposiciones españolas.

p-ingl Preposición o partícula inglesa en general.

Valores: cualquier preposición inglesa o partícula.

parte-de-la-oración Partes de la oración usuales.

Valores: NOUN, VERB, PREP, etc.

prep-espa Preposición española.

Valores: todas las preposiciones españolas.

prep-ingl Preposición inglesa.

Valores: cualquier preposición inglesa.

ptc-ingl Partícula inglesa.

Valores: cualquier partícula inglesa.

tbase1, tbase2 Adverbios españoles o expresiones adverbiales.

Valores: en aquellos casos en que existen, las dos formas posibles del mismo adverbio español.

tipo-foco Tipo de oración utilizado en P-TYPE.

Valores: THATCL y PRPRTCL.

tipo-oración Tipo de una oración (utilizada en C-TYPE).

Valores: Todos los tipos de oración posibles.

tipo-prmod Premodificador de un nombre.

Valores: QUANT, ADJ, NUM y ORD.

tipo-psmod Tipo de modificador utilizado en P-PSMODS.

Valores: INFCL, PP, PRPRTCL, PREDADJ, PTPRTCL and RELCL.

verb-espa Verbo español.

Valores: cualquier verbo español.

Apéndice B. Marcas Semánticas

Como ya se ha señalado, este apéndice contiene parte del trabajo realizado por Teófilo Redondo Pastor como miembro del grupo de PLN del Centro de Investigación UAM - IBM, y sobre la clasificación propuesta se apoya actualmente la definición que se hace de las entradas del diccionario bilingüe.

NOMBRES

Esta clasificación es una aproximación preliminar a la subclasificación semántica para los nombres de la lengua inglesa. La marcas semánticas se han determinado principalmente para nombres individuales.

Figura 55 (Página 1 de 3). Marcas semánticas para nombres		
Descripción	Característica	Ejemplos
<i>entidad abstracta</i>	ABS	industry, market
<i>acción</i>	ACT	analysis, creation
<i>tiempo (edad)</i>	AGE	
<i>actividad (deverbal)</i>	AKTV	writing, dancing
<i>animado</i>	ANIM	any HUM or ANML
<i>ser viviente (animal)</i>	ANML	dog
<i>área</i>	AREA	memory, RAM
<i>actitud</i>	ATTD	patience, approach
<i>edificio</i>	BLDG	laboratory, university
<i>sensaciones físicas</i>	BSNS	hunger, pain, need
<i>circunstancia</i>	CIRC	
<i>conjunto (colección)</i>	CLLT	string
<i>componente (miembro de)</i>	CMP	symbol
<i>computación ()</i>	CMPT	function, command
<i>complejo (derivado de)</i>	CMPX	

Figura 55 (Página 2 de 3). Marcas semánticas para nombres		
Descripción	Característica	Ejemplos
<i>color</i>	COLR	white
<i>causa</i>	COS	
<i>documento</i>	DOC	paper, chapter, appendix
<i>descripción</i>	DSCR	
<i>tiempo (duración)</i>	DUR	day, month
<i>suceso (deverbal)</i>	EVT	jump, processing
<i>mueble</i>	FRNT	chair, wardrobe
<i>tiempo (frecuencia)</i>	FRQ	
<i>característica</i>	FT	
<i>término gramatical</i>	GT	noun, verb, adjective
<i>humano</i>	HUM	user, programmer
<i>información ()</i>	INFO	question, answer
<i>instrumento</i>	INST	
<i>lugar</i>	LOC	stage, street
<i>matemáticas</i>	MATH	even
<i>proceso mental</i>	MEPR	thought
<i>movimiento</i>	MOVE	departure, approach
<i>espacio (medida)</i>	MSRM	inch, millimeter
<i>material</i>	MT	plastic, silicon
<i>número</i>	NUM	one, two, three
<i>opción</i>	OPT	
<i>actor / realizador</i>	PERF	compiler, interpreter
<i>abstracto (filosofía)</i>	PHIL	condition, hypothesis

Figura 55 (Página 3 de 3). Marcas semánticas para nombres		
Descripción	Característica	Ejemplos
<i>objeto físico</i>	PHOB	book, car
<i>espacio (lugar)</i>	PL	
<i>ser viviente (planta)</i>	PLT	rose
<i>nombre de persona</i>	PN	John, Mary
<i>computación (procedimiento)</i>	PROC	function, command
<i>profesión</i>	PROF	soldier, teacher
<i>personalización</i>	PRSL	system
<i>cualidad (deadjetival)</i>	QUA	simplicity, difficulty
<i>ruta (desde)</i>	RFR	
<i>rango / alcance</i>	RNG	proximity, distance, covering
<i>resultado</i>	RSLT	change
<i>ruta (a través)</i>	RTHR	entrypoint
<i>ruta (hacia)</i>	RTO	
<i>espacio ()</i>	SP	
<i>deporte</i>	SPRT	football
<i>estado</i>	ST	approximation
<i>tamaño (deadjetival)</i>	SZ	longitud
<i>método / técnica</i>	TEK	algorithm, approach
<i>tiempo ()</i>	TME	
<i>herramienta</i>	TOOL	
<i>dispositivo / unidad</i>	UD	keyboard, cable
<i>tiempo atmosférico</i>	WTHR	rain, heat, wind

ADJETIVOS

Lo que sigue es una primera aproximación a una subclasificación semántica para los adjetivos ingleses. Se pretende que esta subclasificación sea similar a la propuesta para nombres.

Figura 56 (Página 1 de 2). Marcas semánticas para adjetivos		
Descripción	Característica	Ejemplos
<i>habilidad</i>	ABLE	readable
<i>abstracto</i>	ABST	successful
<i>tiempo (edad)</i>	AGE	old
<i>aspecto (físico)</i>	ASPC	beautiful, dark, light, pale
<i>aptitud</i>	ATTD	ambitious, lucky, patient
<i>sensación física</i>	BSNS	hungry, painful
<i>circunstancia</i>	CIRC	happy (habitualmente TEMP)
<i>color</i>	COLR	blue, red
<i>carácter</i>	CRTR	brave
<i>descriptivo (dinero)</i>	DSMN	cheap, expensive
<i>tiempo (duración)</i>	DUR	daily, monthly
<i>emotivo</i>	EMOT	new, poor
<i>sentimiento</i>	FEEL	happy, pleasant
<i>tiempo (frecuencia)</i>	FRQ	daily, monthly
<i>material</i>	MT	wooden
<i>origen</i>	ORIG	Spanish, Jewish
<i>calidad</i>	QUA	difficult, simple
<i>rango / alcance</i>	RNG	far, near
<i>forma</i>	SHP	circular
<i>sentido</i>	SNS	intelligent, silly
<i>estado</i>	ST	approximate

Figura 58 (Página 2 de 2). Marcas semánticas para adjetivos		
Descripción	Característica	Ejemplos
<i>tamaño</i>	SZ	long, wide
<i>estado temporal</i>	TEMP	empty, full
<i>tiempo atmosférico</i>	WTHR	cold, hot, windy

VERBOS

Los verbos tienen algunas marcas semánticas en PEG, que son las siguientes:

Figura 57. Marcas semánticas de PEG para verbos		
Descripción	Característica	Ejemplos
<i>pregunta</i>	ASKNG	ask, demand
<i>afirmación</i>	ASSRT	affirm, certify, declare, know
<i>causa</i>	CAUSTO	bid, have, make
<i>comunicación</i>	CMMNCT	assure, inform, tell, write
<i>conocimiento</i>	COGN	know, learn, think
<i>sensación</i>	SENS	feel, hear, see
<i>deseo</i>	WSH	hope, wish

Sin embargo, se debe adoptar una clasificación más detallada que permita extraer toda la potencia de un sistema de marcas semánticas. Se propone la siguiente:

Figura 58 (Página 1 de 2). Marcas semánticas para verbos		
Descripción	Característica	Ejemplos
<i>actividad</i>	AKTV	write, dance
<i>agentivo</i>	AGEN	look at, watch, listen to
<i>devenir</i>	BECM	become, end up, get, go, grow, turn
<i>comienzo</i>	BGNN	begin, start
<i>sensaciones corporales</i>	BSNS	ache, hurt, itch, tickle
<i>causativo</i>	CAUSAT	lay, place, put, sit, stand
<i>obligación</i>	CMPUL	compel, constrain, enforce, oblige
<i>continuidad</i>	CONT	continue, go on, keep on

Figura 58 (Página 2 de 2). Marcas semánticas para verbos		
Descripción	Característica	Ejemplos
<i>duración</i>	DUR	last
<i>capacitación</i>	ENABL	enable, entitle
<i>término</i>	END	cease, finish, stop
<i>suceso</i>	EVT	jump, process, destroy, eliminate
<i>gusto / apetencia</i>	LYK	(can't) bear, hate, like, love, prefer
<i>estado mental</i>	MSTA	wonder
<i>movimiento</i>	MOTN	go, move, send
<i>necesidad</i>	NEED	need, require
<i>negación</i>	NEG	mis- , un- , (y otros prefijos negativos)
<i>permiso</i>	PRMS	allow, authorize
<i>suceso (fenómeno)</i>	PHEN	occur, happen, appear
<i>posición</i>	PSTN	situate, finish
<i>rango / alcance</i>	RNG	cover
<i>comunicación (verbos de 'decir')</i>	SAY	say, tell
<i>persuasión</i>	SUAS	agree, decide, intend, suggest
<i>volición</i>	VOLT	want

Apéndice C. Algunos datos de los programas

Todos los procedimientos LISP que se han implementado para MENTOR, tanto las reglas de pretransferencia, transferencia estructural y generación, así como las primitivas del diccionario, se han almacenado de forma independiente y ordenada en un conjunto de librerías LISP. Este tipo de ficheros permite el acceso independiente a cada uno de los procedimientos que contiene, conservando de forma automática información sobre las últimas versiones de cada uno de éstos.

El número de líneas de código de que constan los grupos de funciones indicados es:

Pretransferencia

Total: 571 líneas de código

Transferencia Estructural

- Reglas Estructurales: 2.362 líneas de código
- Utilidades para la Transferencia Estructural: 485 líneas de código

Total: 2.847 líneas de código

Transferencia Léxica

- Primitivas de Decisión: 1.604 líneas de código
- Primitivas de Acción: 683 líneas de código
- Utilidades para la Transferencia Léxica: 758 líneas de código

Total: 3.045 líneas de código

Generación

Total: 2.333 líneas de código

Rutinas de Uso General

Total: 100 líneas de código



Apéndice D. Algunas entradas completas

En este apéndice se presentan algunas entradas completas para dar una visión general de la potencia del uso de las distintas primitivas. Principalmente se presentan verbos, que es el conjunto de elementos más estudiado hasta la fecha, aunque también pueden verse otros elementos de la oración.

Recordemos que este trabajo aún no ha concluido, por lo que las actuales descripciones no se consideran definitivas. Como en el resto de la memoria, se han omitido los símbolos de condición simple, y ésta se señala por la indentación de las líneas.

again ADV

```
+ ((G-FEAT 'POSVB) (G-FLEX '(0 0)) '("otra" "vez"))
```

allow VERB

```
+ ((F-OBJ '(for **)) (G-FLEX '(VERB 0 0)) '("tener" "en" "cuenta"))
+ ((F-OBJ '(of **)) "permitir")
+ ((F-OBJ)
+ ((P-TYPE 'THATCL) "reconocer")
+ ((M-FEAT 'TME) "dejar")
+ ((P-PSMODS 'INFCL) (S-COMPL) "permitir")
+> ("permitir")
+> ("permitir")
```

anybody PRON

```
+ ((P-PSMODS '(RELCL INFCL PTPRTCL PRPRTCL PP)) "cualquiera")
+ ((P-PSMODS 'PREDADJ) "cualquier")
+ ((C-ASPECT 'NEG) "nadie")
+> ("alguien")
```

anything PRON

```
+ ((P-PSMODS '(RELCL INFCL PTPRTCL PRPRTCL PP)
(G-FLEX '(0)) '("cualquier cosa"))
+ ((C-ASPECT 'NEG) (G-FEAT 'SGENDER 'NEUTER) "nada")
+> ((G-FEAT 'SGENDER 'NEUTER) "algo")
```

ask VERB

```
+ ((F-OBJ '(for "")) "pedir")
+ ((F-OBJ '(after "por")) "preguntar")
((F-OBJ)
+ ((P-PSMODS 'INFCL) (S-COMPCL) "pedir")
+ ((F-INDOBJ '(from "")) "pedir")
+ ((M-IS '(question "pregunta")) "hacer")
+> ("preguntar"))
+ ((F-CC 'PP '(about "por")) "preguntar")
+> ("preguntar")
```

be VERB

```
((P-AJP)
((M-FEAT 'WTHR)
+ ((/(& (F-SUBJ) (M-IS '(it "")) (S-XCHANGE NOUNC) "hacer")
+ ((/(& (F-SUBJ) (M-FEAT 'ANIM)) (S-XCHANGE NOUNC) "tener")
+> ("estar"))
+ ((M-FEAT 'STAT) "estar")
+ ((M-FEAT 'TEMP) "estar")
+ ((M-FEAT 'ATTD) (S-XCHANGE NOUNC) "tener")
+ ((M-FEAT 'BSNS) (S-XCHANGE NOUNC) "tener")
+ ((F-CC 'AVP) "estar")
+ ((F-CC 'PP) "estar")
+> ("ser"))
((P-NP)
((P-PSMODS 'PREDADJ)
+ ((M-IS '(old "")) "tener")
+> ((S-XCHANGE PREPC "de") "tener"))
+ ((& (P-PSMODS 'PP 'IN) (P-PRMODS 'NUM) (G-SETPREP "de") "tener"))
+> ("ser"))
+ ((P-PTC '(off "")) (G-FEAT 'CLITIC) "ir")
+ ((P-PTC '(over "")) (G-FEAT 'CLITIC) "terminar")
+ ((F-CC 'PP '(from "de")) "venir")
((F-CC 'PP)
+ ((M-FEAT 'TME) "ser")
+> ("estar"))
+> ("ser")
```

become VERB

```
((P-AJP)
+ ((M-IS 'rich **)) (G-FEAT 'CLITIC) "enriquecer")
+ ((M-IS 'mad **)) "enloquecer")
+ ((M-IS 'ill **)) "enfermar")
+ ((M-IS 'angry **)) (G-FEAT 'CLITIC) "enfadar")
=> ((G-FEAT 'CLITIC) "poner")
((P-NP)
+ ((M-FEAT 'PHOB)
(G-FEAT 'CLITIC) (G-FLEX '(VERB 0)) '("convertir" "en"))
=> ((G-FEAT 'CLITIC) "hacer")
=> ((G-FEAT 'CLITIC) (G-FLEX '(VERB 0)) '("transformar" "en"))
```

benefit VERB

```
+ ((F-OBJ) "beneficiar")
=> ((G-FEAT 'CLITIC) (G-CC 'PP '(from "de")) "beneficiar")
```

by PREP

```
+ ((F-NOFPP 'VERB) **)
=> ("por")
```

calling ADJ

```
+ ((S-RELCL "llamar"))
```

careful ADJ

```
"prudente"
```

easily ADV

```
+ ((G-FEAT 'PSADJ) (S-TBASES '("fácilmente" "con facilidad")))
```


every ADJ

"todo"

everybody PRON

+ ((G-FLEX '(0)) '("todo el mundo"))

function NOUN

"función"

general NOUN

"general"

general ADJ

"general"

general ADV

+ ((G-FEAT 'PREVB)
(S-TBASES '("generalmente" "en términos generales")))

give VERB

((F-OBJ)
+ ((P-PTC '(over **)) (G-FEAT 'CLITIC) "rendir")
+ ((P-PTC '(up **)) (G-FEAT 'CLITIC) "rendir")
=> ("dar")
=> ("ceder")

grow VERB

```

+ ((F-OBJ '(into "")) (G-FLEX '(VERB 0 0)) ('("llegar" "a" "ser")))
((F-OBJ)
+ ((M-FEAT 'PLT) "cultivar")
+ ((G-FEAT 'CLITIC) (G-FLEX '(VERB 0)) ('("dejar" "crecer"))))
+ ((F-CC 'PP '(outof "en")) (G-FEAT 'CLITIC) "originar")
((P-AJP)
+ ((M-IS '(angry "")) (G-FEAT 'CLITIC) "enfadar")
+ ((M-IS '(dark "") (G-FEAT 'CLITIC) "oscurecer")
+ ((M-IS '(old "")) "envejecer")
+ ((M-IS '(cold ""))
((F-SUBJ)
+ ((M-FEAT 'ANIM)
(G-FLEX '(VERB 0 0 0)) ('("empezar" "a" "tener" "frio")))
+ ((M-FEAT 'IMPERSONAL)
(G-FLEX '(VERB 0 0 0)) ('("empezar" "a" "hacer" "frio")))
+> ((G-FEAT 'CLITIC) "enfriar"))
+> ("enfriar")
+> ((G-FEAT 'CLITIC) "poner"))
((F-SUBJ)
+ ((M-FEAT 'PLT) "crecer")
+ ((M-FEAT 'ABS) (G-FEAT 'CLITIC) "desarrollar")
+> ("aumentar"))
+> ("crecer")

```

he PRON

```

+ ((F-IS 'SUBJ) "él")
+ ((F-IS 'OBJ) (G-FEAT 'CLITIC) "le")
((F-IS 'INDOBJ) (G-FEAT 'CLITIC)
+ ((F-IS 'PP) "él")
+ ((& (F-OBJ) (M-SEG 'PRON)) "se")
+> ("le"))
+> ("él")

```

himself PRON

```

+ ((F-IS 'OBJ) (G-FEAT 'CLITIC) "se")
+ ((F-IS 'INDOBJ) (G-FEAT 'CLITIC) "se")
+> ((G-FLEX '(0 0)) ('("él" "mismo")))

```

hold VERB

```

+ ((F-OBJ '(with "con")) (G-FLEX '(VERB 0 0)) ("estar" "de" "acuerdo"))
+ ((F-OBJ '(back "")) (G-FEAT 'CLITIC) "guardar")
+ ((F-OBJ '(down "")) "sujetar")
+ ((F-OBJ '(in "")) "contener")
+ ((F-OBJ '(off "")) "rechazar")
+ ((F-OBJ '(out "")) "extender")
+ ((F-OBJ)
+ ((M-FEAT 'PHOB) "sostener")
+ ((M-IS '(breath "aliento")) "contener")
+ ((F-SUBJ)
+ ((M-FEAT 'CLLT) "contener")
+> ("retener"))
+> ("retener"))
+ ((P-PTC '(back "")) (G-FEAT 'CLITIC) "refrenar")
+ ((P-PTC '(off ""))
+ (G-FEAT 'CLITIC) (G-FLEX '(VERB 0 0)) ("mantener" "a" "distancia"))
+ ((F-SUBJ)
+ ((M-FEAT 'PHIL) "valer")
+> ("resistir"))
+> ("resistir")

```

hot ADJ

"caliente"

hypothesis NOUN

"hipótesis"

In PREP

```

+ ((F-NOFPP 'VERB) (S-INFNP) "al")
  ((F-NOFPP 'NOUN)
+ ((M-IS '(FASHION "moda")) "a")
+ ((M-IS '(HUNDRED "ciento")) "a")
  ((M-IS '(HALF "mitad"))
+ ((M-MORPH 'SING) "por")
  ("en"))
+ ((M-FEAT 'TME)
+ ((M-IS '(DAYTIME "día")) "durante")
+ ((C-ASPECT '(PRES NEG)) "desde hace")
+ ((C-ASPECT '(PAST NEG)) "durante")
+ ((C-ASPECT '(FUTURE NEG)) "durante")
+ ((C-ASPECT '(PAST)) "tras")
+ ((C-ASPECT '(FUTURE)) "dentro de")
  ("en"))
+ ((M-FEAT 'WTHR) "bajo")
+ ((M-FEAT 'MSRM) "de")
+ ((M-FEAT 'COLR) "de")
+ ((& (F-MODHEAD 'NOUN) (P-PRMODS 'ADJ) (M-MORPH 'SUPR) "de"))
+ ((& (M-FEAT 'NUM) (F-MODHEAD 'NOUN) (M-FEAT 'NUM) "de cada"))
+ ((& (P-PRMODS 'NUM) (F-MODHEAD 'NOUN) (M-FEAT 'NUM) "de cada"))
  ("en"))
+ ("en")

```

keep VERB

```
+ ((F-OBJ '(at "de")) (G-FEAT 'NEG) "cesar")
+ ((F-OBJ '(from "de")) (G-FEAT 'CLITIC) "guardar")
+ ((F-OBJ '(back "")) "guardar")
+ ((F-OBJ '(off "")) (G-FLEX '(VERB 0 0)) ("cerrar" "el" "paso" "a"))
+ ((F-OBJ '(on ""))
+ ((P-PRPTCL) "seguir")
+> ("continuar"))
+ ((F-OBJ '(out "")) "excluir")
+ ((F-OBJ '(up "")) "conservar")
+ ((F-OBJ)
+ ((M-FEAT 'PHOB) "guardar")
+ ((P-PSMOODS 'PREDADJ) "tener")
+ ((P-PSMOODS 'PP 'FROM)
+ ((P-PSMOODS 'PRPTCL) (S-COMPCL) (G-SETPREP "") "impedir")
+> ((S-FCHANGE '((PREPC INDOBJ))) "ocultar"))
+ ((P-TYPE 'PRPTCL) "seguir")
+ ((P-PTC '(away ""))
+ (G-FLEX '(VERB 0 0)) ("mantener" "a" "distancia"))
+> ((G-FEAT 'CLITIC) (S-FCHANGE '((OBJ PREPC "con")) "quedar"))
+ ((F-CC 'PP '(to "en")) (G-FEAT 'CLITIC) "mantener")
+ ((F-CC 'PP '(outof "fuera de")) (G-FEAT 'CLITIC) "quedar")
+ ((P-PTC '(away ""))
+ (G-FEAT 'CLITIC) (G-FLEX '(VERB 0 0)) ("mantener" "a" "distancia"))
+ ((P-PTC '(back "")) (G-FEAT 'CLITIC) "apartar")
+ ((P-PTC '(off ""))
+ (G-FEAT 'CLITIC) (G-FLEX '(VERB 0 0)) ("mantener" "a" "distancia"))
+ ((P-PTC '(on "")) "continuar")
+ ((P-PTC '(up "atrás")) (G-FEAT '(CLITIC NEG)) "quedar")
+ ((F-SUBJ)
+ ((M-FEAT 'PLT) (G-FEAT 'CLITIC) "conservar")
+ ((M-FEAT 'ANIM) (G-FEAT 'CLITIC) "conservar")
+> ("conservar"))
+> ("mantener")
```

length NOUN

"longitud"

like VERB

```
((F-OBJ)
+ ((C-ASPECT '(PASSIVE)) (G-FEAT 'NOPRON) "gustar")
+= ((S-FCHANGE '((OBJ SUBJ) (SUBJ INDOBJ))) "gustar"))
+= ("gustar")
```

long ADJ

```
+ ((& (F-NOFADJ) (M-IS '(TIME "tiempo")) (G-FEAT 'PRENOUN) "mucho"))
=> ("largo")
```

look VERB

```
+ ((F-OBJ '(at "")) "mirar")
+ ((F-OBJ '(after "")) "cuidar")
+ ((F-OBJ '(for "")) "buscar")
+ ((F-OBJ '(up "")) "buscar")
=> ("parecer")
```

make VERB

```
+ ((F-OBJ '(out "")) "distinguir")
+ ((F-OBJ '(up ""))
  ((C-ASPECT '(PASSIVE))
   (G-FEAT 'NONACTIVE) (G-STATE) (G-CC 'PP '(of "por")) "componer")
=> ("constituir"))
+ ((F-OBJ)
  ((P-NP)
   ((M-FEAT 'HUM) "nombrar")
   ((S-FCHANGE '({OBJ PREPC "de"})) "hacer"))
  ((P-PSMODS 'PREDADJ)
   ((M-IS '(ready "")) "preparar")
   "hacer"))
+ ((P-PSMODS 'INFCL) (S-INFNP) (G-FLEX '(VERB 0)) '("obligar" "a"))
+ ((M-FEAT 'MONEY) "ganar")
+ ((M-IS '(mistake "error")) "cometer")
+ ((F-CC 'PP '(into "en")) (G-FEAT 'CLITIC) "convertir")
=> ((G-FEAT 'NONACTIVE) (G-STATE) "hacer"))
+ ((F-CC 'PP '(after "")) "seguir")
+ ((F-CC 'PP '(for "a"))
  ((M-FEAT 'PL) (G-FEAT 'CLITIC) "dirigir")
  "contribuir"))
+ ((F-CC 'PP '(towards "a")) (G-FEAT 'CLITIC) "dirigir")
+ ((P-PTC '(away "")) (G-FEAT 'CLITIC) "escapar")
+ ((P-PTC '(off "")) (G-FEAT 'CLITIC) "escapar")
+ ((P-PTC '(out "")) (G-FLEX '(VERB 0)) '("salir" "bien"))
+ ((P-PTC '(up "")) (G-CC 'PP '(for "")) "compensar")
=> ("hacer")
```

missing ADJ

+ ((S-RELCL "faltar"))

morning NOUN

+ ((F-IS 'PP '(IN "por")) (G-FEAT 'SGENDER 'FEMALE) "mañana")
 => ("mañana")

move VERB

((F-OBJ)
 + ((M-FEAT 'HUM) "trasladar")
 + ((P-PTC '("aside "")) "apartar")
 + ((P-PTC '("away "")) "quitar")
 + ((P-PTC '("back "hacia atrás")) "mover")
 + ((P-PTC '("down "")) "bajar")
 + ((P-PTC '("forward "hacia adelante")) "mover")
 + ((P-PTC '("up "")) "subir")
 => ("mover"))
 + ((F-CC 'PP '("into "a")) (G-FEAT 'CLITIC) "trasladar")
 + ((F-CC 'PP '("to "a")) (G-FEAT 'CLITIC) "trasladar")
 + ((F-CC 'PP '("off "de")) (G-FEAT 'CLITIC) "alejar")
 + ((F-CC 'PP '("out "")) (G-FEAT 'CLITIC) "salir")
 + ((F-CC 'PP '("outof "de")) (G-FEAT 'CLITIC) "salir")
 + ((P-PTC '("about "")) (G-FLEX '("VERB 0)) '("dar" "vueltas"))
 + ((P-PTC '("aside "")) (G-FEAT 'CLITIC) "apartar")
 + ((P-PTC '("away "")) (G-CC 'PP '("from "de")) (G-FEAT 'CLITIC) "alejar")
 + ((P-PTC '("back "")) "retroceder")
 + ((P-PTC '("forward "")) "avanzar")
 + ((P-PTC '("in "")) (G-CC 'PP '("to "en")) (G-FEAT 'CLITIC) "instalar")
 + ((P-PTC '("on "")) "seguir")
 + ((P-PTC '("out "")) (G-FEAT 'CLITIC) "salir")
 => ((G-FEAT 'CLITIC) "mover")

nothing PRON

+ ((G-FEAT 'SGENDER 'NEUTER) "nada")

nothing NOUN

```
((P-PRMODS 'ADJ)
+ ((M-IS '(mere "")) "bagatela")
+ ((M-IS '(sweet "")) "terneza")
+> ("naderia"))
+> ("nada")
```

nothing ADV

```
((G-FLEX '(0)) '("de ningún modo"))
```

nowhere NOUN

```
((G-FLEX '(0)) '("ninguna parte"))
```

nowhere ADV

```
((C-ASPECT 'STAT) (G-FLEX '(0)) '("en ninguna parte"))
+> ((G-FLEX '(0)) '("a ninguna parte"))
```

old ADJ

"viejo"

people NOUN

```
((P-PRMODS 'NUM) "persona")
+> ((G-FEAT 'SNUMBER 'SING) "gente")
```


play VERB

```
((F-OBJ)
+ ((M-IS '(part "cometido")) "desempeñar")
+ ((M-FEAT 'SPRT) (S-FCHANGE '((OBJ PREPC "a"))) "jugar")
+ ((M-FEAT 'HUM) (S-FCHANGE '((OBJ PREPC "con"))) "jugar")
=> ("tocar")
((F-CC 'PP)
+ ((M-FEAT 'INST) "jugar")
=> ("tocar")
=> ("tocar"))
```

return NOUN

"retorno"

return ADJ

```
+ ((G-FLEX '(Ø Ø)) '("de" "retorno"))
```

return VERB

```
+ ((F-OBJ) "devolver")
+ ((F-CC 'PP '(to "a")) "regresar")
=> ("volver")
```

run VERB

```
((F-OBJ)
+ ((P-PTC '(down ""))
  (G-FLEX '(VERB 0 0 0))'("interrumpir" "el" "funcionamiento" "de"))
+ ((M-FEAT 'CMPT) "ejecutar")
=> ("correr"))
+ ((P-PTC '(down "")) (G-FLEX '(VERB 0 0 0))'("dejar" "de" "funcionar"))
+ ((P-PTC '(out "")) (G-FEAT 'CLITIC) "acabar")
+ ((F-CC 'PP '(on "con")) "funcionar")
+ ((F-CC 'PP '(outof "sin")) (G-FEAT 'CLITIC) "quedar")
((P-AJP)
+ ((M-IS '(dry "")) (G-FEAT 'CLITIC) "secar")
  (F-SUBJ)
+ ((M-FEAT 'CMPT) (G-FEAT 'CLITIC) "ejecutar")
+ ((M-FEAT 'UD) "funcionar")
=> ("correr"))
=> ("correr"))
((F-SUBJ)
+ ((M-FEAT 'CMPT) (G-FEAT 'CLITIC) "ejecutar")
+ ((M-FEAT 'UD) "funcionar")
=> ("correr"))
=> ("correr"))
```

something PRON

"algo"

something NOUN

"algo"

something ADV

"algo"

stop VERB

```
((F-OBJ)
+ ((P-TYPE 'PRPTCL) (S-INFNP) (G-FLEX '(VERB 0)) ('dejar' "de"))
+ ((P-PSMODS 'PRPTCL) (S-COMPCL) "impedir")
  ((P-PSMODS 'PP 'FROM)
+ ((P-PSMODS 'PRPTCL) (S-COMPCL) (G-SETPREP "") "impedir")
=> ("parar"))
=> ("parar"))
+ ((F-CC 'PP '(by ""))
  (G-FLEX '(VERB 0 0 0)) ('hacer' "una" "breve" "visita"))
=> ((G-FEAT 'CLITIC) "parar"))
```

that ADJ

"ese"

that ADV

"tan"

that CONJ

"que"

that PRON

```
+ ((M-FEAT 'DEMO) "ése")
+ ((F-IS 'PP) (G-FLEX '(ADJ 0)) ('el' "que"))
=> ("que")
```

think VERB

```
+ ((F-OBJ '(of "en")) "pensar")
=> ("pensar")
```

to PREP

```
((F-NOFPP 'NOUN)
+ ((M-FEAT 'TME) "hasta")
  ((& (F-MODHEAD 'NOUN)
    ((M-FEAT 'PROF) "de")
  ))
+ ((F-IS 'PP '(FROM "desde")) "hasta")
+ ((P-PRMODS 'QUANT "por")))
((M-FEAT 'PL)
+ ((C-ASPECT '(MOVE)) "a")
+> ("hasta"))
+> ("a"))
+> ("a")
```

want VERB

```
((F-OBJ)
+ ((P-PSMODS 'INFCL) (S-COMPCL) "querer")
+> ("querer"))
+> ("querer")
```

what PRON

```
((C-TYPE 'QUES) (G-FEAT 'SNUMBER 'SING) "qué")
+> ("lo que")
```

where CONJ

"donde"

where PRON

```
((G-FLEX '(0 ADJ 0)) '("en" "el" "que"))
```

where ADV

"dónde"

which ADJ

"qué"

which PRON

+ ((F-IS 'PP) (G-FLEX '(ADJ 0)) '("el" "que"))
+> ("que")

whom PRON

+ ((C-TYPE 'QUES) (G-FLEX '(0)) '("a quién"))
+ ((F-IS 'PP) (G-FLEX '(ADJ 0)) '("el" "que"))
+> ((G-FLEX '(0 PRON)) '("a" "quien"))

yourself PRON

+ ((F-IS 'OBJ) (G-FEAT 'CLITIC) "se")
+ ((F-IS 'INDOBJ) (G-FEAT 'CLITIC) "se")
+> ((G-FLEX '(0 0)) '("usted" "mismo"))